DEVELOPING ROBOTS THAT IMPACT HUMAN-ROBOT TRUST IN EMERGENCY EVACUATIONS

A Thesis

Presented to

The Academic Faculty

by

Paul Robinette

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in

Robotics

Georgia Institute of Technology

December 2015

DEVELOPING ROBOTS THAT IMPACT HUMAN-ROBOT TRUST IN EMERGENCY EVACUATIONS

Approved by:

Ayanna M. Howard, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Karen M. Feigh
School of Aerospace Engineering
Georgia Institute of Technology

Alan R. Wagner, Co-Advisor
Aerospace, Transportation and Advanced Systems Laboratory
Georgia Tech Research Institute

Andrea L. Thomaz
School of Interactive Computing
Georgia Institute of Technology

Henrik I. Christensen
School of Interactive Computing
Georgia Institute of Technology

Approved on November 4, 2015

Acknowledgements

So many people have supported me during my graduate work and contributed to my research that it will be impossible to thank them all in this space. My wife deserves the first position for encouraging me when I needed it the most. Thank you, Priety, for reminding me why I decided to go to graduate school, for keeping me focused when I needed to be working and for distracting me when I needed to take a break.

My advisors, Dr. Ayanna Howard and Dr. Alan Wagner, have provided the years of good guidance necessary for this work. This thesis topic started as a discussion between Dr. Howard and me during my first week at Georgia Tech. During the discussion, we realized that the field of human-robot interaction was focused on humans providing instructions to robots, but had little work on robots providing instructions to humans. After many talks with Dr. Wagner, we moved the focus from simple compliance with robot instructions to a more fruitful focus on human-robot trust in this domain. This dissertation would not have been completed without help from both of them. I am truly thankful to have two advisors who care enough to discuss every aspect of this work with me.

My committee, Dr. Henrik Christensen, Dr. Karen Feigh, and Dr. Andrea Thomaz, also deserve thanks. Each has been willing to meet with me to provide advice and assistance whenever I needed it from the very beginning of this work. Much of this work started as projects in their courses and it would not have been possible without their support.

Thanks to Larry Labbe, the Georgia Tech Fire Marshal, and the Georgia Tech Fire Safety Office for their advice, assistance and equipment in running the final experiment. That experiment would not have been possible without their help. Thanks to Wenchen Li and Robert Allen for their help in setting up and running the final experiment.

Every person in The HumAnS lab has helped me during the course of this work by aiding me during experiments, being guinea pigs in pilot studies, and by relaxing together during much needed time off. Thank you to Stephen Williams, Douglas Brooks, Lonnie Parker, Chung Hyuk Park,

Gregorio Drayer, Hae Won Park, Rick Coogle, Sergio Garcia, LaVonda Brown, Kevin DeMarco, Brittney English, Anthony Spears, Nicole Giullian, and Mason Nixon. Special thanks to Hae Won for reading my work and providing encouragement as needed.

The RoboGrads organization has been very helpful throughout my time at Georgia Tech. Everyone involved in the organization deserves thanks for making the program what it is today. Michael Novitzky has always been there to help me whenever I needed it. Paul Drews and I have worked together many times over the last decade (or more) and I have enjoyed every minute. Thanks also to Kelsey Hawkins and Brian Hrolenok. Thanks to Tucker Hermans and Brian Goldfain for organizing so many evenings out. My quals study group, Baris Akgun, Martin Levihn, Ana Huamán, and Phil Grice, helped me through the qualifying exams and gave me much needed advice throughout the program. Thanks to Jake Huckaby and Tiffany Chen for preceding me on the Executive Board and making my time as president easier.

Thank you to Dr. Mick West for hiring me at GTRI and to Dr. Tom Collins for being one of the best project managers I have every worked with. Thank you to all of my other coworkers as well, but especially David Jensen, Dr. Charles Pippin, Dr. Zsolt Kira, Andrew Price, and Andrew Melim.

Of course, my family deserves thanks as well. My parents have always encouraged my educational pursuits. They provided me with the discipline to finish what I started. My sister has always been willing to talk when I needed some help. Everyone in my family has been supportive, even if they did not always understand why I would want to spend so much time in school.

Many people contributed to my decision to go to graduate school in the first place. In particular, Ryan Meuth provided me with motivation and advice since the earliest days of my academic career. Beth Meuth has always been there for myself and my wife, as well. Thank you to Ryanne Dolan for working with me on so many academic and professional projects over the years. Thank you to Joe Siebert for all of our discussions over the years. Thank you to Mark Herrera for providing perspective at times. Thank you to Nicholas Lessley for convincing me to take some time off for relaxation at some times and for providing advice at others.

Table of Contents

List of Tables

List of Figures

Summary

High-risk, time-critical situations require trust for humans to interact with other agents even if they have never interacted with the agents before. In the near future, robots will perform tasks to help people in such situations, thus robots must understand why a person makes a trust decision in order to effectively aid the person. High casualty rates in several emergency evacuations motivate our use of this scenario as an example of a high-risk, time-critical situation. Emergency guidance robots can be stored inside of buildings then activated to search for victims and guide evacuees to safety. In this dissertation, we determined the conditions under which evacuees would be likely to trust a robot in an emergency evacuation.

We began by examining reports of real-world evacuations and considering how guidance robots can best help. We performed two simulations of evacuations and learned that robots could be helpful as long as at least 30% of evacuees trusted their guidance instructions. We then developed several methods for a robot to communicate directional information to evacuees. After performing three rounds of evaluation using virtually, remotely and physically present robots, we concluded that robots should communicate directional information by gesturing with two arms. Next, we studied the effect of situational risk and the robot's previous performance on a participant's decision to use the robot during an interaction. We found that higher risk scenarios caused participants to align their self-reported trust with their decisions in a trust situation. We also discovered that trust in a robot drops after a single error when interaction occurs in a virtual environment. After an exploratory study in trust repair, we have learned that a robot can repair broken trust during the emergency by apologizing for its prior mistake or giving additional information relevant to the situation. Apologizing immediately after the error had no effect.

Robots have the potential to save lives in emergency scenarios, but could have an equally disastrous effect if participants overtrust them. To explore this concept, we created a virtual environment of an office as well as a real-world simulation of an emergency evacuation. In both, participants interacted with a robot during a non-emergency phase to experience its behavior and then chose whether

to follow the robot's instructions during an emergency phase or not. In the virtual environment, the emergency was communicated through text, but in the real-world simulation, artificial smoke and fire alarms were used to increase the urgency of the situation. In our virtual environment, we confirmed our previous results that prior robot behavior affected whether participants would trust the robot or not. To our surprise, all participants followed the robot in the real-world simulation of an emergency, despite half observing the same robot perform poorly in a navigation guidance task just minutes before. We performed additional exploratory studies investigating different failure modes. Even when the robot pointed to a dark room with no discernible exit the majority of people did not choose to exit the way they entered.

The conclusions of this dissertation are based on the results of fifteen experiments with a total of 2,168 participants (2,071 participants in virtual or remote studies conducted over the internet and 97 participants in physical studies on campus). We have found that most human evacuees will trust an emergency guidance robot that uses understandable information conveyance modalities and exhibits efficient guidance behavior in an evacuation scenario. In interactions with a virtual robot, this trust can be lost because of a single error made by the robot, but a similar effect was not found with real-world robots. This dissertation presents data indicating that victims in emergency situations may overtrust a robot, even when they have recently witnessed the robot malfunction. This work thus demonstrates concerns which are important to both the HRI and rescue robot communities.

Chapter 1

Introduction

One day soon, robots will interact with people in everyday activities. It is safe to assume that these robots will perform many everyday tasks, such as delivering packages, running errands, and cleaning houses. Robots will likely also take on many responsibilities in high-risk domains, such as assisting medical staff in hospitals, fighting fires, and rescuing victims in emergencies. Each of these applications requires people to trust the robots that they depend on to complete tasks. It is possible that some people will undertrust robots, by expecting them to not perform tasks they are designed to be capable of completing. It is also possible and, as we show, likely that people will overtrust robots by expecting them to perform tasks that they might not be capable of completing successfully. Overtrusting robots can lead to dangerous situations where people put themselves at unnecessary risk by believing robots will mitigate that risk.

It is especially important for people to properly trust a robot in high-risk, time-critical situations. A fire emergency is one such scenario. A fire emergency requires people to evacuate a building quickly. Evacuees have little time to make decisions or select optimal paths, so they rely on existing technology, such as emergency exit signs and evacuation maps, as well as information gleaned from authority figures, to find the best way out. As robots become more pervasive in everyday life, we can expect them to one-day guide evacuees during emergencies. There is considerable risk of injury or even death to evacuees in this situation, so we must understand the factors that affect human-robot trust in these scenarios before such robots are deployed.

This dissertation addresses human-robot trust as it applies to humans accepting guidance from autonomous robots during a high-risk, time critical situation such as an emergency evacuation. The goal of this work is to develop a robot that is capable of guiding evacuees during an emergency

and, in doing so, determine the level of trust people will place in this robot. Thus, we can study human-robot trust decisions as a component of a robot that saves lives. In pursuit of this effort, we have developed a model of evacuee behavior, created methods for robots to guide humans, studied human-robot trust as affected by robot performance and situational risk, validated the emergency guidance robot platform, and explored methods to repair trust after an error.

1.1 Motivation

Trust is a requirement in every interaction that involves risk, from everyday tasks to life-and-death situations [30, 78]. Victims in emergencies do not waste precious time arguing with firefighters and other emergency responders; they follow the responders' directions because they trust the information provided by the agent. In much the same way, store owners trust their cashiers to handle money during transactions in order to do business. The risk involved in these two scenarios is different, but the concept of trust is inherent in both.

Robots have incredible potential to assist humans in everyday and emergency tasks. One such task is aiding victims during a fire. Concerned about high casualty rates in emergency situations such as the Station Nightclub Fire of 2003 [32], we have explored numerous situations where emergency guidance robots can improve human survivability in evacuations [59, 60, 61]. Research that examined high-rise evacuations indicates that these buildings are especially difficult to evacuate quickly [27]. As the number of high-rise buildings continues to increase, we expect the need for additional emergency guidance technologies to also increase [64]. Robots represent one type of new technology that can save lives in emergencies.

Today, robots are being actively deployed in scenarios that range from cleaning floors to bomb disposal; however such tasks either present low risk to humans (e.g. cleaning a floor) or are tightly controlled by human experts (e.g. bomb disposal). To increase the potential for autonomous robots to aid people in additional high-risk tasks, people must first trust the robots to perform these tasks correctly. Moreover, we must investigate the conditions under which people will overtrust or undertrust robots in these tasks.

1.2 Virtual, Remote, and Physical Presence Experiments

Throughout this dissertation, we present experiments that use virtual, remote, or physical presence of the robot. There are many different factors that influence which presence level is best for a particular

human-robot interaction study. Below, we briefly discuss the advantages and disadvantages of each.

1.2.1 Physical Presence Experiments

A physical human-robot interaction experiment requires the use of an actual robot (as opposed to a virtual robot) and thus typically requires physical space to perform the experiment [6]. The physical space is most often a laboratory, but can also be a house, a public place or a workspace such as a factory or office. Regardless, the participant and/or the robot, along with any other necessary equipment, must be transported to the location of the experiment. Transporting a robot can be expensive and prone to errors. Robots used in experiments are typically under active development and thus are often unsuited for locations far from the laboratory. Convincing participants to come to a laboratory to perform an experiment can also be costly and results in self-selection: only those who have spare time and means of transportation are likely to participate. For this reason, many HRI experiments performed in university laboratories utilize students of the university as participants (for example, [6, 24]).

Many HRI experiments are appropriate to administer in a laboratory setting. For example, teaching by demonstration typically requires participants to touch or closely observe a physical robot and does not depend on its surroundings in any particular way. Other HRI experiments, such as those involving search and rescue robots, present problems for experimenters. It is difficult to transform a laboratory into a believable disaster area. Previous work has presented experiments with a selection of props and a written scenario [47]. Others use specially built areas, such as the Disaster City at Texas A&M (see [26] and [50] for examples), but such areas are rare and expensive to create.

Most laboratory robots are under active development and thus are not completely free of errors. An error made by a robot during an experiment can potentially injure a participant and will almost certainly affect the response of the participant. Another potentially confounding factor is noise and other distractions from nearby laboratories as an experiment is in progress. Even the presence of the experimenter can affect the outcome of the experiment. Controlling these factors in a laboratory setting requires considerable effort. The main advantage of performing an experiment with a physical presence robot is that the participant experiences every aspect of the actual robot in question. Many components of a robot cannot be simulated accurately, so it is often necessary to perform a physical experiment in order to test the complete system.

1.2.2 Virtual Presence Experiments

We define a virtual human-robot interaction experiment as an experiment where participants observe and interact with a simulation of a robot through a computer. Others have referred to this as a Virtual Environment (VE) experiment and, when combined with additional simulation hardware, this paradigm has been called an Immersive Virtual Environment (IVE) experiment [10, 76]. The robot must be entirely simulated and the interaction must take place in some sort of a virtual environment, similar to interactions in video games. This paradigm is attractive because most scenarios that are difficult to create in a laboratory are fairly easy to create using modern three-dimensional modeling software and game engines. It is possible to create the exact scenario that the experimenter would like to test in a virtual environment. This paradigm has been used in numerous social psychology experiments (see [10] for examples).

Another benefit of virtual experiments is that they can be deployed to participants anywhere in the world via the internet. Most game engines have an option to create a web-based game that can be loaded by a web browser plugin. Even participants with no video game experience can then interact with a virtual robot in the environment chosen by the experimenter. Recently, many experiments have used this technique to increase the number of participants who experience their robot [51, 55, 62]. Crowdsourcing an experiment on the internet using services such as Amazon's Mechanical Turk allows for a larger participant population base than is typically available for physical experiments. Other studies have found that Mechanical Turk provides a more diverse participant base than traditional human studies performed with university students [53, 15, 8, 37]. These studies found that the Mechanical Turk user base is generally younger in age but otherwise demographically similar to the general population of the United States (at the time of those studies, Mechanical Turk was only available in USA). Crowdsourcing also allows the experiment to be performed in parallel with typically much faster results than physical experiments. A virtual experiment that requires one hundred participants to each spend one minute interacting with a robot can have final results in minutes or hours, rather than the days or weeks necessary to recruit, assemble, and supervise such a population for a physical experiment. Thus, virtual experiments are most useful in situations where the experimenter wishes to iterate through prototypes or pilot studies rapidly.

The behavior of a simulated, virtual robot can be controlled easier than a physical robot. This is not applicable for user studies or other studies where the quirks of the robot are being examined, but can be helpful when exploring the behaviors a robot should perform to effectively influence a human participant. As an example, consider a study that measures the loss of trust a participant

experiences in a robot after the robot performs specific errors. If the robot performed other errors than those specified in the experimental case, or malfunctioned during the control case, then the results would have to be discarded and additional participants would be required. By using a simulation environment, we can tightly control the behavior of the robot and ensure that no unintentional errors were committed.

Virtual experiments are not without their problems. Participants must volunteer for the experiment, thus there is still a self-selection bias in the participant population. This is balanced by allowing a much larger body of participants to volunteer through the use of the internet. While virtual experiments remove the possibility of noise and other distractions from nearby laboratories, they lose the ability to tightly control the environment in which a participant performs the experiment. A participant may choose to perform the experiment while watching television or listening to music and thus miss an important component. This can be mitigated by asking participants to explain their responses, thus ensuring that a thoughtful process was used in their actions, and by asking participants questions which check their understanding of the experiment. Additionally, other studies have found that social interactions between humans and robots are not always well represented through non-physical presence [6, 79].

1.2.3 Remote Presence Experiments

The use of video streaming technology for remote presence experiments allows for a happy medium between virtual and physical experiments. In remote experiments (or Video Human-Robot Interaction experiments [76]), participants view a video of a robot (either prerecorded or live) and complete their tasks through a web interface. Remote experiments allow participants to observe the actual robot hardware as it performs its experimental tasks, but do not allow participants to touch the robot. The use of prerecorded videos allows experimenters to gather participant feedback on designs that are still under active development and might not perform perfectly in every trial. Additionally, videos can be recorded or streamed from a laboratory setting, which allows participants to be involved in the experiment even if they cannot physically travel to the laboratory. Remote experiments can often be crowdsourced, similar to virtual experiments. Videos can be placed on a service like Amazon's Mechanical Turk and be available to a larger participant population than physical experiments. These experiments have similar drawbacks to virtual experiments, with the one major improvement being the use of the actual robot in the experiment to remove any effect simulation artifacts would have on participant responses. The remote presence paradigm has been previously

tested in [6, 79].

1.3 Contributions

Throughout the course of this work, we have found the following thesis statement to be true:

> *Most human evacuees will trust an emergency guidance robot during an evacuation if they understand the robot's instructions and it exhibits efficient guidance behavior.*

The converse statement, that evacuees will not trust a robot that does not provide efficient guidance, has been found to be true in virtual studies, but not in physical studies. We have conducted thirteen human subject experiments with a total of 2,168 participant (2,071 in virtual or remote studies conducted over the internet and 97 in physical studies on campus) to support this thesis statement. Participants ranged in age from 18 to 72, had educational backgrounds ranging from less than a high school diploma to a Ph.D. and diverse occupations. Throughout these studies, we have produced five major contributions to the field.

To begin this work, we had to understand how people behave in evacuations and determine the potential effect, if any, of robots in emergency situations. Using research from the fields of psychology and sociology as well as after disaster reports, we developed a model of evacuee behavior in an emergency. We then simulated this model under emergency conditions with and without robots to find the effect of robots on these situations. This research is presented in Chapter 3 and led to our first contribution:

> *1. Developed a model of group affinity and information propagation between evacuees in emergency situations and evaluated the model with automated evacuation guides.*

These early results provided evidence that robots could be useful as guides in an emergency. Next, we developed methods for a robot to communicate guidance information during an emergency. An experiment, discussed in Chapter 4, demonstrated that participants quickly lost trust in a robot when the robot was unable to communicate clearly. In Chapter 5, we present several information conveyance modalities and their evaluations with human participants. This led to our second contribution:

> *2. Developed models for communicating directional information to humans in high-risk, time-critical situations and identified their correlation to various robot form factors.*

Our models were shown to be effective at providing understandable guidance, but it was not yet known if people would trust a robot in this task. Wagner's definition of situational trust (presented

in [78] and discussed in terms of this work in Section 2.3) tells us that trustees make their decision to trust or not based on the amount of risk they are facing in the situation and their model of the trustee's performance. Thus, in Chapter 6, we present experiments that test the effect of situational risk and variable robot performance on an individual's propensity to trust a robot. This led to our third contribution:

3. Measured the effect of risk modality and robot effectiveness on human-robot trust.

Next, we validated the emergency guidance robot system in virtual and real-world experiments (Chapter 7). We began by asking participants to choose to follow robots or emergency exit signs in a timed virtual experiment. We then developed a more complex virtual experiment that allowed participants to explore some of their environment and experience the robot in a single interaction before a simulated emergency began. After these two experiments confirmed that our robots worked and that participant's responded in a similar way as in previous work, we developed a real-world experiment that allowed participants to experience the actual robot in a simulated emergency. While this experiment did not confirm our previous findings, it did lead to our fourth contribution:

4. Measured a person's propensity to follow an emergency guidance robots in a realistic emergency scenario.

Under some conditions, robots will almost certainly break trust in real-world scenarios. Thus, we used existing techniques identified in psychological and sociological literature to give the robot the ability to repair trust after a mistake. Using our virtual office evacuation simulator, where we previously developed behaviors capable of breaking trust, we verified that some of these techniques successfully repaired broken trust in an emergency scenario (Chapter 8). This led to our final contribution:

5. Developed techniques to repair broken trust between a human and a robot.

1.4 Scope

This work has many applications to human trust in social robots, but this dissertation is focused on developing robot behaviors to guide humans in short-term, high-risk situations, similar to the circumstances in an emergency. We assume that humans in an emergency situation will not have considerable prior experience with guidance robots, so our experiments have been designed for participants using our robots for the first time.

Our work treats trust as a binary decision: either the person trusts the robot or the person does not. In time-critical situations there is typically only time to make one decision to follow or not follow a navigation aid. Additionally, our work is focused on Wagner's definition of trust because we are only concerned with the concept of a human putting his or her outcomes at risk, dependent on the actions of a robotic agent. This is discussed in greater detail in the next chapter.

1.5 Dissertation Outline

In the next chapter we discuss other work related to the topics found in this dissertation. Following that, in Chapter 3 we present our model of evacuee behavior and two simulations which tested it. In Chapter 4, we develop two prototype robots and present a pilot study to evaluate them. Chapter 5 builds on this by describing various methods a robot can use to guide an evacuee and evaluates them. Chapter 6 presents experiments that determined the factors that affect a person's decision to trust a robot in an emergency. In Chapter 7, we validate all of the work so far in realistic emergency situations. Chapter 8 then presents techniques a robot can use to repair broken trust. Finally, in Chapter 9, we conclude this work and give recommendations for future directions of research in this field. Table 1.1 lists the experiments performed for this dissertation and their locations in the text.

Table 1.1: Experiments described in this dissertation and their major findings.

Experiment Title	Location in Text	Major Findings	Number of Participants
Group Affinity Simulation	Section 3.2	Robots can improve survivability rates in emergencies.	0 (simulated humans)
Information Propagation Simulation	Section 3.3	At least 30% of evacuees need to trust robots to improve survivability.	0 (simulated humans)
Robot Prototypes Evaluation	Chapter 4	Participants followed robots as long as they understood the instructions.	15
Virtual Presence Robot Instructions	Section 5.2	Participants could understand a dynamic sign when close to the camera and multiple arms performing gestures at both distances tested.	208
Remote Presence Robot Instructions	Section 5.3	Validated virtual results in the remote paradigm.	128
Physical Presence Robot Instructions	Section 5.3	Validated virtual and remote results in a physical experiment paradigm.	48
Trust Narratives	Section 6.2	Determined the extent that Wagner's trust definition corresponded with participant trust definitions.	180
Single Round Robot Guidance - Bonus Motivation	Section 6.2	Iteratively developed evacuation simulation.	210
Single Round Robot Guidance - Emergency Motivation	Section 6.2	Validated Wagner's definition of trust with emergency guidance robots.	120
Double Round Robot Guidance - Bonus Motivation	Section 6.3	Discovered that monetary bonus did not provide sufficient motivation for participants to make a deliberate choice.	126
Double Round Robot Guidance - Emergency Motivation	Section 6.3	Determined the effect of robot performance on human-robot trust in emergencies.	129
Robots vs. Existing Guidance Technology	Section 7.2	Found that participants noticed robots more often than emergency exit signs in a simulated emergency.	111
Virtual Office Evacuation	Section 7.3	Reaffirmed our previous findings that prior robot performance affects human-robot trust in simulated emergencies.	114
Physical Office Evacuation	Section 7.4	Participants trusted robots despite prior behavior in physical evacuation simulation.	49
Trust Repair	Chapter 8	Found that timing of trust repair techniques had significant effect on restoring trust.	730

Chapter 2

Related Work

We begin our discussion of related work by describing existing emergency guidance technology and its impact on our robot design. We then discuss human behavior in emergency scenarios. Following this, we explain our conception of trust (following Wagner's work [78]) and then list related studies in the field of human-robot trust. We conclude with several studies of human-robot interaction that use similar methodologies to our own.

2.1 Existing Emergency Guidance Technology

To facilitate an orderly evacuation, robot guides should be designed such that it is obvious they are pointing evacuees towards an exit. Inspiration for our robots was taken from several studies on exit sign design. A NIST report confirmed previous studies which found that luminosity is a large factor in visibility (some of their observers suggested it is the largest factor) [21]. No recommendation could be made about other factors involved in an exit sign, as some observers preferred one particular style and others preferred another. The color red was preferred to green, however the authors mention that this could be due to familiarity with the color or differing brightness levels. We use red in our designs because our robots are designed for use in North America where red exit signs are popular.

Another study evaluated several different exit signs in use at the time, but did not reach many conclusions despite testing in normal conditions and smokey conditions [57]. They determined that color, brightness, and size of the sign mattered, but could only recommend that signs be as large and bright as possible. They found that exit signs in North America were usually red while those in Europe were usually green. Green signs typically allow for a greater luminosity, but easy recognition is also important, so the study could not give a firm recommendation on color.

Exit signs also must consider people with disabilities. This is an area where robots could be of great benefit as they can approach those who have vision problems. A study was performed where people in assisted living facilities rated the visibility of various exit signs [12]. Many of their participants had vision problems. This paper had some surprising results as it shows that people with a vision disability can recognize an exit sign at about the same distance as those without vision disabilities (mean of 13.9-14.6 meters, depending on the sign, for those with disabilities, 14.5-14.7 meters otherwise). The study found that people can recognize an exit sign at a point several meters past where they can read the word "EXIT". In Chapters 4 and 5 we present guidance robots that use elements of the designs recommended by the above studies.

2.2 Human Behavior in Emergencies

Several studies have been performed on how people react in specific emergency situations [32, 27, 71, 14]. A report analyzed the 1993 World Trade Center bombing and found that occupants of the two towers were generally reluctant to exit in a timely fashion [27]. Many preferred to stay on their floor awaiting further instructions instead of evacuating, even after fire alarms sounded. Occupants were reportedly surprised that they had to wait for several hours for firefighters to arrive on their floor to provide further instructions. This further motivates the need for automated assistance in emergencies and presents evidence that people do not always take fire alarms seriously.

In related after-disaster research, Helbing et al. analyzed video of crowds panicking during the 2006 Hajj in Mecca, Saudi Arabia [34]. The researchers plotted the position and velocity of each person in the area immediately in front of a bridge entrance. From this, they determined when the crowd transitioned from laminar to stop-and-go or turbulent flows. Using this data, they made several recommendations to the Saudi Arabian government to improve the flow of pedestrians and reduce the number of casualties. These recommendations included making certain pathways one-way, discouraging stops on walkways, and tracking the number of people in each area. This work was a continuation of past research that defined a social force model to describe human walking movements [35, 38]. Other researchers have noticed differences in walking behavior between cultures and developed a model sensitive to that variable [28]. This body of work shows the importance of affecting crowd behavior early, while it is still in a laminar flow mode, in order to avoid injuries.

One study interviewed 128 survivors from a fire in the Solarium of the Summerland Leisure Complex in 1973 [71]. The researcher found that individuals with strong ties to a group were less likely to panic and try to escape in a selfish way than previously thought. They found that families

and groups of friends were more likely to make escape choices that were better for the group as a whole. Sometimes, particularly tight groups would make escape choices that benefitted the group despite great risk to the individual making the choice. For example, parents tend to refuse to leave a burning building without their children. This study showed that some families that were not together at the onset of the emergency still found each other and exited as a group. The affiliate behavior was greatly dependent on the closeness of the group. Families were much more likely to stay together, close friends somewhat less and casual acquaintances (such as those who met at the resort) were unlikely to stay together at all. We develop a computational model of evacuee behavior based on this research in Section 3.2.

The previous research has shown how people in groups move through crowds in an emergency, but another study experimented with what exit individuals chose in a simulated emergency [7]. The researcher recruited volunteers for a simulated emergency situation at an IKEA store. Each volunteer was given a headset which played an alarm and gave instructions to evacuate as quickly as possible. The study found that when volunteers could see closed exit doors nearby they preferred to go out through the front of the store, but when they could see an open exit door (such that they could see outdoors) then they were more likely to exit through the open door regardless of distance. This shows the difficulty in convincing people to follow evacuation plans with existing technology.

Several experiments have been performed to determine the best way to evacuate airplanes during emergencies. During one test, researchers identified the impact of aisle width on time to evacuate through over-wing exits [49]. They determined that wider aisles (up to approximately 20 inches) allowed more people to evacuate. Greater than 20 inches of width and the aisle became wider than the exit itself, so evacuees assumed that more than one person could leave at a time. This was not possible due to the width of the exit door, so this caused a bottleneck in the exit row. The researchers also examined what happened when volunteers were given extra incentive to evacuate quickly. This incentive was an additional $7.75 over their pay as volunteers if they could be among the first 50% to evacuate. For over-wing exits, this actually increased the mean time for evacuation. Some volunteers pushed through bottlenecks to get out faster, which delayed the group as a whole. Volunteers also climbed over seats (the researchers note that participants climbed over seats occupied by other participants at the risk of injuring both parties) to jump ahead in the line. Researchers have also examined passenger confusion about the operation of exit doors [72]. These studies were helpful in determining the motivations to give to participants in our experiments. The studies showed that a small amount of money can produce highly motivated behavior in participants.

Several simulators, similar to those we present in Chapter 3, have been created to help design

buildings and new technology to reduce evacuation time. The ESCAPES simulator was used to model the Los Angeles airport in order to aid security efforts to streamline emergency procedures [73, 74]. Another simulator modeled the Station Nightclub fire of 2003 to help researchers understand which factors contributed to the high casualty rate [19]. A computational model of evacuation that studied the destructive forces of a crowd has been created, but not evaluated in reference to real-world emergencies yet [52]. These simulators tend to be either proprietary or designed for a single scenario, so we were not successful in using any in our research, despite our best efforts to contact the researchers who created them. Instead, we examined the literature created about each simulator to inform our own simulator designs in Chapter 3.

2.3 Conceptualizing Trust

Numerous researchers have proposed conceptions of trust that range from computational implementations of cognitive processes [17], to neurological changes in reciprocity games [42], to a probability distribution over an agent's actions [30]. Other researchers consider trust to have multiple forms, depending on the actors and environment [36]. After a review of the available literature, Lee and See conclude that trust is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability [44]. Building from Lee and See's definition of trust Wagner states that trust is "a belief, held by the trustor, that the trustee will act in a manner that mitigates the trustor's risk in a situation in which the trustor has put its outcomes at risk" [78].

2.3.1 Representing Interactions

Outcome matrices are a useful tool for formally conceptualizing social interaction. These matrices (or normal-form games in the game theory community) explicitly represent the individuals interacting as well as the actions they are deliberating over. The impact of each pair of actions chosen by the individuals is represented as a scalar number or outcome. For interactions involving trust, it is common to label one individual as a trustor and the other as a trustee. An example using the Investor-Trustee game (see [42] for an example) is presented in Figure 2.1. The investor (or trustor) has the choice of investing $10 with the trustee or not. If the investor does decide to invest, the investment triples according to the rules of the game and the trustee can decide how much to return. In this simplified version of the game, the trustee only has two options: return an even split or take all of the money. As this example shows, outcome matrices can be used to create a simple, numerical representation of a social interaction where risk can be analyzed with ease.

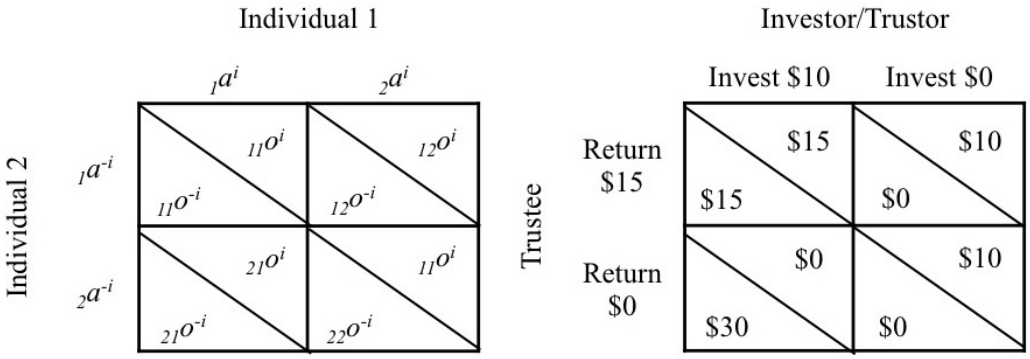


Figure 2.1: An example outcome matrix is depicted formally and as an investment game. The risk associated with the trustee's action can be approximated by subtracting the values on the right in the invest $10 columns.

2.3.2 Conditions for Situational Trust

The outcome matrix representation can be used to formally represent the types of situations that have long been used in trust research [40]. The investment game, for example, presents an investor with some amount of money. The investor must decide whether to invest the money with a trustee or not. If the investor chooses to invest, the money invested appreciates to some larger amount. The trustee must then decide what amount to return to the investor. Figure 2.1 depicts an example game. Investment games such as this have become the de facto method for investigating trust by the trust research community [42]. While we continue to use this trust game as an example in this section, we instead choose to use emergency evacuations as our trust situation for this dissertation because of the heightened risk involved in a choice that could affect a person's survival.

In the matrix presented in Figure 2.1, for example, the investor has a choice: she can choose to invest or not to invest. Likewise, the trustee can choose to return some amount of money or not to return any money. Although the matrix is a specific example of a situation involving trust, we can easily abstract away the actions that the actors are deliberating over to develop a series of conditions for trust. These conditions (see Figure 2.2), which are derived from our working definition for trust, can be used to segregate outcome matrices into those that require trust on the part of a trustor and those that do not [77].

2.4 Human-Robot Trust

Much of the current research in human-robot trust can trace its roots to human-automation research over the last few decades. Muir developed a model for the trust a person places in a machine based

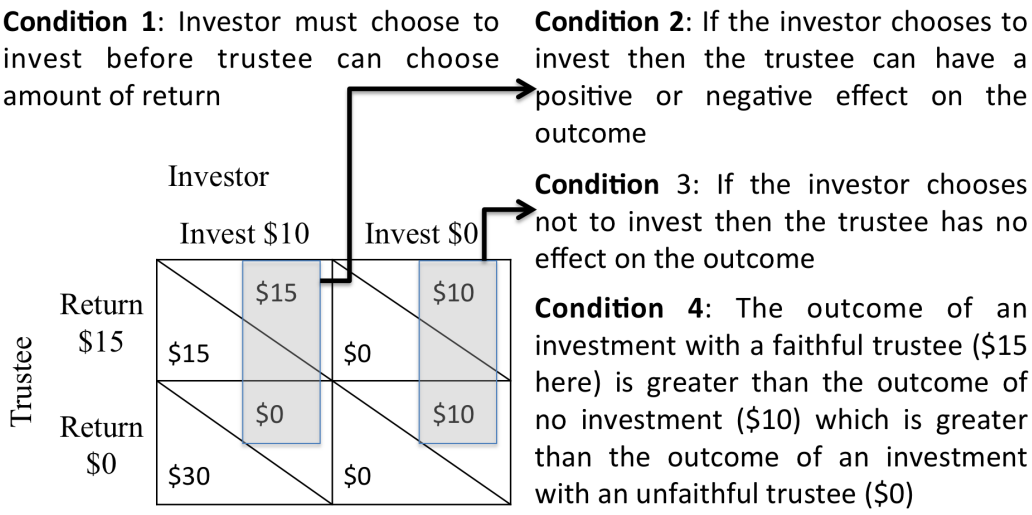


Figure 2.2: The conditions for trust derived from Wagner's definition for trust are shown above with examples from the Investor-Trustee game.

on sociological definitions of trust [48]. The model states that trust is a set of expectations that the machine will be technically competent, responsible, and will act in good faith.

In related human-automation research, Parasuraman and Riley identify reasons that operators use automation or not in certain situations [54]. They identify many positive uses, where the automation is used as designed by operators, but also describe situations where operators misused or abused automation and where operators did not use technology when it could have been beneficial. In general, operators used automation technology in appropriate situations when they were well-informed about its benefits and risks, but over-relied on it when overloaded or poorly trained. Likewise, operators did not use automation when it raised too many false alarms or when they perceived the likelihood of error (caused by operator or machine) as higher with the machine than without. The authors make an important note that, at some level, trust in an automated system actually means trust in the designer of the automated system. As such, a robot that has been designed by a top research institute to perform a guidance task in an emergency might be more trustworthy in that situation than a robot without those features, regardless of the robot's prior performance.

Lyons and Stokes continue this field of research with a study that showed participants in their experiment relied on automation technology more as the risk of the scenario increased, even though the real-world incident (involving an airplane collision) that motivated their study was caused by under-use of automation [45]. This demonstrates that people will tend to trust automation, such as a robot, more if they perceive the risk of their situation to be high.

In a review of existing human-automation trust research, Hoff and Bashir propose a three layered model of trust: the disposition of the operator/participant to trust the system, the situational factors that affect trust (such as the risk), and the learned factors (such as prior robot performance). They recommend that automated technology use anthropomorphic qualities and exhibit polite behavior to engender trust. They found that greater transparency and ease of use also increased participant trust in the experiments they surveyed.

A survey of human-robot trust research found that prior robot performance had the greatest impact on trust in the robot [33]. Other attributes of the robot, such as its appearance, had a small correlation to trust. This motivated our use of robot performance to bias participants for or against trusting the robot in Chapters 6 and 7. Environmental factors, such as the task performed in the experiment, were also found to have an effect on participant trust, which we further investigate in Section 6.3.

Following this work, Yagoda and Gillan developed a scale to measure the amount of trust in a human-robot system [82]. The scale gives several statements about user, sensor, effector, and automation reliability and predictability, each of which are rated on a Likert scale. As described in Section 2.3, our definition of trust is concerned with one agent putting his or her outcomes at risk in the hands of another. As such, we prefer to record participant actions to determine if participants trust our robot, rather than give participants a questionnaire about system reliability.

Related research has focused on the factors that participants indicate affect their trust in a robot [16]. Carlson et al. finds that reliability and reputation impact trust in surveys of how people view robots. Again, in contrast to surveys, we use immersive simulations and real-world experiments to record a person's actual behavior during an interaction involving trust. We also focus on initial interactions with a robot, rather than trust that has been built over a long history. In additional research testing the relationship between trust and a robot's performance, van der Brule et al. presented two experiments that asked participants to observe videos of a robot or interact with a virtual robot that exhibited two different levels of motion fluency and two different levels of performance [76]. The authors found that performance had a significant effect on participant trust levels, but that motion fluency did not affect trust in the robot when participants interacted with the robot (as opposed to when they only observed it).

Desai et al. performed several experiments related to human-robot trust which are well-regarded by the human-robot interaction community [23, 24, 39, 25, 22]. This group performed experiments on two different university campuses to determine the effect of robot reliability on an operator's decision to trust the robot. Participants were given the option to interrupt autonomous operation of

the robot and take manual control at any time during a search and rescue task. The experimenters measured the number of incorrect paths taken by the robot, the number of obstacles hit by the robot, and the number of victims recorded by the operator. Participants were rewarded with a larger compensation for a better performance. The researchers found that trust in the robot was decreased by poor robot performance, but that trust could be regained if the robot performed better in later parts of the experiment. In a later experiment, they found that if the robot warned operators that it would have reduced reliability in the near future then operators would take manual control without losing their trust in the robot. In contrast to the work by Desai et al., our work and the emergency evacuation scenario we investigate does not afford an opportunity for the human to take control of the robot. Instead, we are examining situations when people must choose to either follow the guidance of a robot or not. While this still explores the level of trust a person is willing to place in an autonomous robot, we believe that the difference between an operator's perspective on trust and an evacuee's perspective on trust is significant. The evacuee cannot affect the robot in any way and must choose between his or her own intuition and the robot's instructions.

Mason et al. use a maze environment to explore the impact of robot reliability on participant decisions to follow the robot [46]. Many of the study's findings, however, are inconclusive. Although their work bears some conceptual similarities, the research we present here is focused on investigating the impact of trust on a person's decision-making during high-risk situations, such as emergency evacuation.

Researchers have also examined a human's decision to follow a robot's directions. Bainbridge et al. found that participants were likely to follow odd and potentially destructive instructions from a robot under certain conditions [6, 5]. Our research does not examine odd or destructive instructions, but does investigate the factors that influence a person's decision to follow instructions from a robot in an emergency situation. Their work indicates that participants are predisposed to following instructions from a robot, even though their study involved little or no perceived risk to the participants themselves.

Salem et al. performed an experiment to determine the effect of robot errors on unusual requests [65]. They found that participants still completed the odd request made by the robot in spite of errors. In their work, robot errors consisted of navigation malfunctions and incorrectly following a request from the participant, but the requests did not necessarily involve any of the those same functions of the robot. Thus, participants could have believed that the robot's navigational system was malfunctioning, but that it was otherwise competent. In our work, we strive to have the robot commit errors that are clearly related to functions that participants would depend on in a trust

situation so that we can attempt to bias participants for or against trusting the robot.

2.5 Human-Robot Interaction

Considerable research has focused on using robots for search and rescue applications. Bethel and Murphy studied how volunteers reacted to rescue robots in a simulated urban disaster [9, 50]. They created several recommendations for how robots should approach, contact, and interact with the victims. For the approach and other motions, the researchers suggest using smooth acceleration and deceleration. In contrast, robots are usually jerky when moving in an unknown environment. The researchers also suggested using blue lighting around the robot to convey a sense of calm. For interaction, they note that there are several different "zones" where the robot can be: the intimate zone (0 to 0.46 meters), the personal zone (0.46 to 1.22 meters), the social zone (1.22 to 3.66 meters) and the public zone (further than 3.66 meters). Robots are assumed to stay in the social zone or closer. To communicate, the researchers assumed that the robots would have to be in the intimate or personal zones. They suggested using voice communication to reassure the victim and music when there is no information to communicate. More recent work has extended this to aerial vehicles [26]. This body of work provides guidelines for where an emergency guidance robot should be placed to communicate effectively.

Simulated emotions have also been tested to see how they can improve human responses when a robot instructs a human to leave a room due to an unexpected emergency [47]. This work began by using videos posted online to determine if humans could understand the emotions being displayed by the robot [55]. The robot gave clear, verbal instructions aided by emotional actions, so participants were only tested on their ability to understand the robot's emotional actions and comply with its requests. We used a similar approach starting with crowdsourcing and ending with laboratory studies to test robot understandability and refine our guidance robot designs in Chapter 5 . Studies in non-verbal robot communications have found that robots and humans work better in teams when the robot performs non-verbal cues and gestures during the interaction [13]. This inspired our work in Chapter 5 to develop non-verbal methods of communicating information across a distance to evacuees.

Orkin and Roy were inspired by early chatbots, such as ELIZA, to create a game to simulate interactions between two people in a restaurant by crowdsourcing on the internet [51]. Participants would join the game and randomly be assigned as either a waiter or a client. Then they would proceed to interact as if they were in an actual restaurant. The researchers noted that participants

typically took the game seriously and acted as if they were in a real social situation. The experiment generated considerable data related to responses to typical prompts in the environment. Participants were solicited through blogs, web postings, emails and social media. A total of 3,355 participants played 5,200 games over several months and completed a survey afterwards. Other research has expanded on this crowdsourced data gathering process to help train a robot for a space mission [20]. The simulation involves two participants on a simulated Mars base, one as the robot and one as the astronaut.

Other researchers have explored the difference between various robot-presence paradigms in different research areas. Related work has found that participants are generally unlikely to follow "unusual requests," such as throwing textbooks in the trash, given by a remote presence robot as compared to a physically present robot [6]. Other work has found that physically present robots are rated better than their virtual or remote counterparts at coaching tasks [79]. In [70], robots were more effective in influencing human participants for 3D tasks, but virtual agents were more effective for 2D tasks. Another study found that virtual agents and physical agents both had their benefits and problems when conducting discussions with participants about health topics [56]. This gives evidence that robots must be present to have a social effect on human participants in real-world tasks. A study that compared responses to videos of a robot approaching a human actor with real-world responses to a robot approaching a participant found little difference in the methodologies [80].

2.6 Robots in Emergency Evacuations

In previous evacuation robot research, robots with directional audio beacons [68] were deployed in optimal positions to reach as many people as possible [69]. These robots were shown to decrease the total amount of time to evacuate in a simulation of an emergency. Physical robots were also deployed in a building to show that the system can automatically redeploy due to the loss of a robot. This research focused on using the robots primarily as static beacons to attract attention to the best exit. The placement of robots was tested in an experiment where some robots failed in their mission, but a human subject experiment was not performed. We expand on this idea by determining the level of trust a participant would place in such a robot, as well as by developing visual methods of communication for emergency guidance robots.

Following our initial work on emergency guidance robots (Chapters 3 and 4), others have developed simulators to test different algorithms for robots in emergency scenarios [83, 11]. Zhang and Guo used potential fields to model the physical environment during a fire emergency and simulated

evacuees in a manner similar to ours in Section 3.2 [83]. Boukas et al. developed a new crowd modeling technique using cellular automata and tested it by having a robot provide guidance to human participants in a simulated evacuation [11]. Participant trust in the robot was not evaluated and the authors did not report motivating participants to find an exit quickly. Instead, the authors used this experiment to validate their prior simulations on crowd movement. Additionally, other researchers have created a 3D simulator for evaluating human-robot trust in emergency scenarios, similar to those we present in Chapters 6 and 7, but have not used the simulator in a human subject experiment [3]. This group has previously performed work in a social interface for human-machine trust [4], trust in human-robot teamwork [2], and using simulated emotions to engender trust [1].

As stated above, significant work has been performed by other researchers to understand the motivations of evacuees in emergencies. We use this work to make computational models of evacuees in Chapter 3 so that we can evaluate the effect of robots in large-scale evacuations. Other researchers have also developed robots that aid emergency personnel in rescuing people after a disaster scenario. Instead, our work focuses on providing guidance to evacuees before a building collapses, in the hope that fewer victims will need to be rescued later. Chapters 4 and 5 focus on our development of these robots. To this point, the field of human-robot trust has focused on the trust an operator places in a robot. We use this work as a starting point, but our work tests the level of trust a person will place in a robot when they have no option to control the robot. Chapters 6 through 8 explore this topic.

Chapter 3

Evacuee Behavior

3.1 Introduction

If we are to study the decisions made by humans in short-term, high-risk situations while interacting with a robot then we must first understand the behavior of humans in these situations without robots. Emergency evacuation is a problem in this domain that is very well studied. For example, shouting "FIRE" in a crowded movie theater can cause many injuries and even deaths from the subsequent rush to the exits. In a real fire, with smoke and alarms, the chaos is even worse. Even simulated emergencies can cause injuries to volunteers [49, 72]. Also, consider two evacuations, both caused by fires, that are particularly well documented: the 1973 Summerland Leisure Complex Fire and the 2003 Station Nightclub Fire. To understand the behavior of humans in high-risk, time-critical situations, we created a computational model of group behavior during evacuations and a model of information propagation among evacuees using reports from these two disasters. Based on these reports and our own simulations, we have identified critical tasks that robots could perform to aid human evacuation. This leads to the first contribution of this work:

> *Developed a model of group affinity and information propagation between evacuees in emergency situations and evaluated the model with automated evacuation guides.*

3.2 Group Affinity in Emergencies

In this section, we created a model of human evacuation behavior ranging from high to low affinity groups based on work discussed in Chapter 2 [71]. We then utilized this model to evaluate the effect on an evacuation when a guide robot assists.

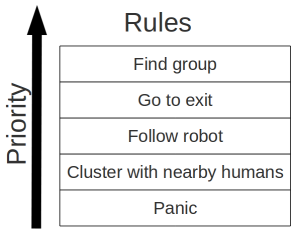


Figure 3.1: High affinity rule priorities

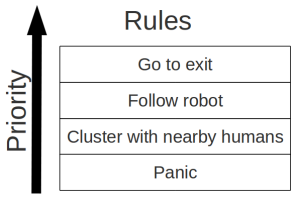


Figure 3.2: Low affinity rule priorities

3.2.1 Model of Human Evacuation Behavior

Rules for the human evacuation model were defined such that humans would find other humans, move towards exits or exit signs and follow evacuation guidance robots. Rules had priorities (see Figures 3.1 and 3.2) based on which rule people are likely to follow in each situation. The highest priority rule that could be executed was followed in each situation.

High affinity groups (such as families) first search for other members of their own group before attempting to exit. This rule superseded all other rules, including those that would allow the individual to exit sooner. The next rule defined how humans assembled as a group, regardless of affinity. Lower affinity groups use this as a first rule, while higher affinity groups would follow this rule only after their group was assembled. This rule is similar to actual behaviors during emergencies where people tend to crowd together in hopes of finding an exit. This behavior also means that as humans tend to move toward an exit, other humans who cannot see the exit but who can see the humans leading the group will tend to move towards safety.

The model suggests that humans will follow a robot as soon as the robot is seen. We assume that people will treat the robot as a mobile exit sign and head directly towards it while trying to maintain some group cohesion with nearby humans and family groups. Similarly, the model has a rule that humans will proceed directly to an exit as soon as it is seen. High affinity groups make sure that others in their group are likely to see (and thus exit) before they exit themselves. This assumes that the humans know that the exit leads directly outside (see [7]).

The lowest priority rule for the model is that a human would panic. This rule would only be

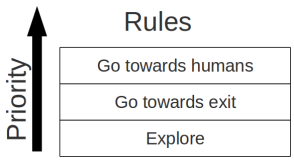


Figure 3.3: Evacuation robot rule priorities

executed when no humans, exits, or robots are near enough to be seen. This rule models how people move randomly (if at all) in a panic. Placing it at the lowest priority is justified by assertions in related work that humans do not actually panic in an emergency [71].

3.2.2 Evacuation Guidance Robot Behavior

A rule-based behavior was created for the evacuation robots with priorities applicable in each situation. The robot's first priority is to search for as many evacuees as possible. To accomplish this, it moves towards nearby humans. Once the robot reaches the humans, it begins to head toward the nearest exit. If the robot moves too far from its group of evacuees, it goes back for humans that might not be able to see it any more. Once the humans start to exit, the robot starts to move towards the back of the group of humans so that it can guide those who may not be able to see the exit yet. Once all humans within sensor range have been evacuated, the robot explores to find more evacuees. The priorities for these rules can be seen in Figure 3.3.

3.2.3 Experimental Setup

As a first simulation, we wanted to test the effect of emergency guidance robots on high and low affinity groups moving around a large room. Humans and robots were each simulated with a rule-based planner and simple trajectory planning. Each entity was allowed to move one unit of distance (0.14 meters) each iteration (0.1 seconds). Four exits were placed in the 500 by 500 unit (approximately 70 meter by 70 meter) environment, one at each corner. The simulation was run for 1000 iterations (enough time for an entity to cross the width of the environment twice, 100 seconds). Two versions of the simulation were run: one with robots and one without. The percentage of humans evacuated in this time was used as a metric.

3.2.3.1 Human Simulations

Humans were assumed to have full 360 degree awareness of objects near them. They had a sight range of 100 units (15 meters) to see lighted objects, such as robots and exit signs, but only 50 units

(7.5 meters) for other humans. This was based on research that suggests people can recognize a lighted exit sign from a much greater distance than they could recognize another object [12]. A total of 100 simulated humans were split into 20 groups with random sizes in a Gaussian distribution with mean of five and standard deviation of two individuals. Each group was given a random affinity value between zero and one. The groups were placed in the environment on a Gaussian distribution centered at a random point with a 50 unit (7.5 meter) standard deviation in each dimension so that most members of the group would be visible to most other members. Each human's behavior depended on their group affinity, the proximity of humans around them, visible exit signs, and nearby robots.

High Affinity Groups Groups were randomly classified as close-knit groups, such as families. These humans' highest priority was to find the other members of their group. As a first choice, group members would proceed to the average position of the other members of their group, regardless of their distance. They would ignore all other humans, robots and exits. It was assumed that the family was able to communicate over larger distances than they could see. Once the group was together, they would look for an exit. If an exit was within sight then they would proceed to that exit. If not, they would look for a robot within sight. If they could find a robot, they would proceed towards that robot as a group. Humans averaged the center of mass of all visible humans with the center of their family group and the robot's position to determine their new goal position. If no robot was found, the human would average their family's center point with the center of mass of visible humans and head to that spot. This made human evacuees group together. If no humans or robots were in range, the human would panic and move randomly. The method to determine a high affinity group member's goal can be seen in Algorithm 3.1.

Low Affinity Groups Individuals randomly assigned to the low affinity condition ignored their group. If an exit was available, they would proceed directly towards it. If a robot was visible, they would proceed towards the average of the center of mass of any visible humans and the position of the robot. This produced a line of humans following the robot. If no robots were nearby, the human would proceed to the center point of the visible humans. Again, if no humans or robots were nearby then the human would panic and move randomly. The method to determine a low affinity group member's current goal can be seen in Algorithm 3.2.

Algorithm 3.1 High affinity group guidance

```
groupCentroid = average(all members of group)
humanCentroid =
     average(all humans within 50 units)
if dist(groupCentroid, myPosition) > 50:
     goal = groupCentroid
else if dist(nearestExit, myPosition) < 100:
     goal = nearestExit
else if dist(nearestRobot, myPosition) < 100:
     goal = average(nearestRobot, humanCentroid,
          groupCentroid)
else if dist(nearestHuman, myPosition) < 50:
     goal = average(groupCentroid, humanCentroid)
else:
     goal = randomPoint
```

Algorithm 3.2 Low affinity group guidance

```
humanCentroid =
     average(all humans within 50 units)
if dist(nearestExit, myPosition) < 100:
     goal = nearestExit
else if dist(nearestRobot, myPosition) < 100:
     goal = average(nearestRobot, humanCentroid)
else if dist(nearestHuman, myPosition) < 50:
     goal = humanCentroid
else:
     goal = randomPoint
```

Algorithm 3.3 Robot guidance

```
goal = nearestExit
humanCentroid =
     average(all humans within 100 units)
if 50 < dist(humanCentroid, myPosition) < 100:
     goal = humanCentroid
else if dist(goal, myPosition) < 50:
     goal = randomPoint
```

3.2.3.2 Robot Simulations

Robots were assumed to have sensors that could detect any humans within 100 units (15 meters) in any direction and were given knowledge of the position of each exit. Four robots were used for the experiment. They were placed at positions towards the center of the environment. Robots were given an initial goal of their nearest exit. If no humans were within sensing range, the robot would proceed towards that goal. If humans were within sensing range, the robot would proceed towards the center of all visible humans. Once the center of these humans was close (within 50 units, 7.5 meters), the robot proceeded towards the nearest exit. If the center point of humans drifted outside of the 50 unit range, the robot would turn back to gather the evacuees together again. Once the robot reached the exit, it would wait for all humans to exit and then head to a random point in the environment. If it intercepted humans along the way, it would start over and guide them to the nearest exit. The function to define the robot's current goal can be seen in Algorithm 3.3.

3.2.4 Results

Twenty runs of this experiment were performed. Figure 3.4 shows the mean percentage of humans evacuated in 1000 iterations. A one tailed T-Test was performed with $t(18)=24.8$, $p < 0.01$.

Robots helped to evacuate many more people than trials without robots; however, the tested human models do not allow for much exploration, so if they cannot see a robot or an exit sign they panic. Some interesting behaviors emerged during the simulation, illustrated in the examples in Figure 3.5. In Figure 3.5a, the robots and humans can be seen at their starting positions. Robots are shown as red circles. Exits are shown as red squares. Humans of the same group are given the same color.

In Figure 3.5b the humans can be seen converging on the robots. The interesting point to note

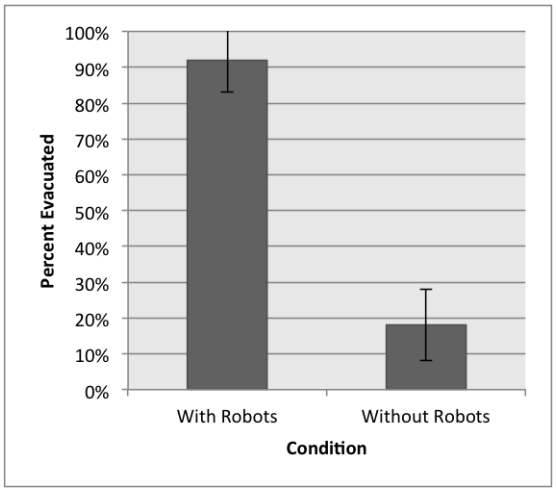
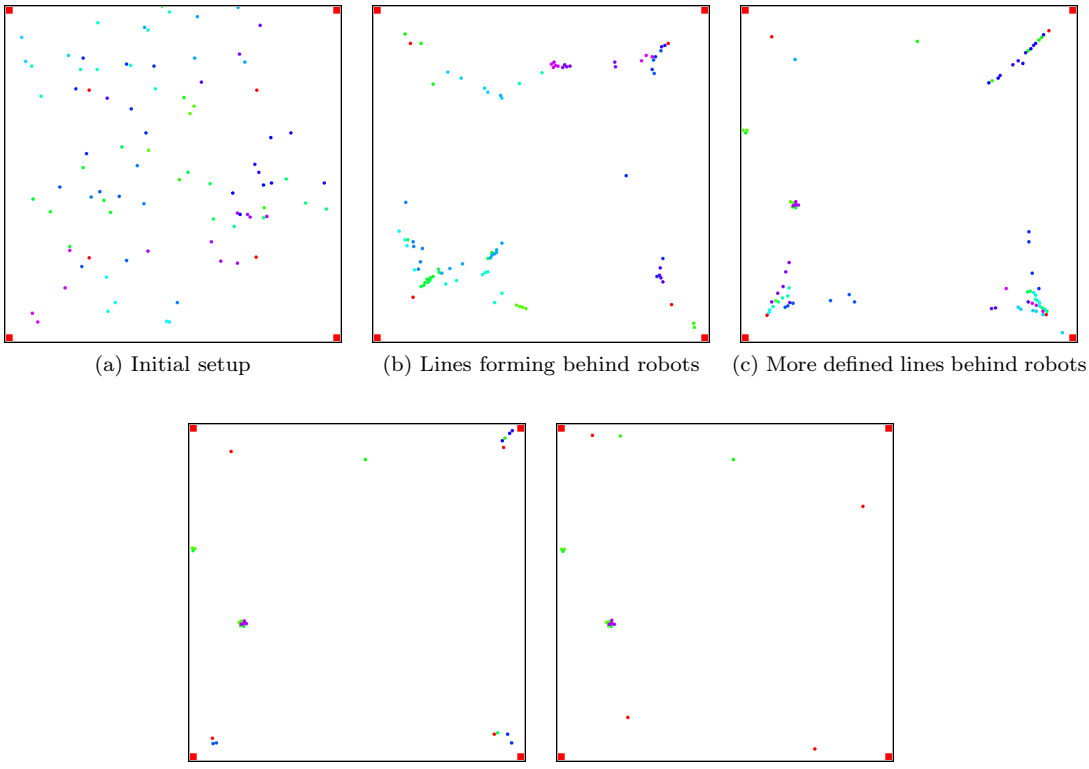


Figure 3.4: Results of evacuation simulations with and without robots. Error bars represent standard deviation.



(a) Initial setup (b) Lines forming behind robots (c) More defined lines behind robots

(d) Robots waiting for humans to exit (e) Robots looking for more humans

Figure 3.5: Example visualizations from emergency evacuation simulations. Robots are represented as red circles, humans are represented as other circles (color-coded by group) and exits are shown as red squares in the corners.

here is that the humans towards the end of the line cannot see the robot, they are simply following the group in front of them. This shows that, according to the current model, not every person in a group needs to be directly led, some will follow others in the hopes that they are heading to safety.

In Figure 3.5c, a relatively straight line can be seen in the top right. Note that this was from a different run than the previous figure, but it is typical. A narrow 'V' can be seen in the lower left. The lower right is more disjoint because the group was less orderly to start with. At this point, the humans are in the process of exiting. Note the two groups along the left who have missed the robots and the one straggler along the top.

In Figure 3.5d, most of the humans have exited and the robots are waiting for the last humans in their immediate area to leave. The robots will oscillate at this step as they move toward stragglers at the end of the line and then guide them towards the exit.

Finally, in Figure 3.5e, the robots are exploring again to look for survivors. The robot in the top left has found a straggler and is guiding him or her to the nearest exit.

3.2.5 Discussion

Using a simulation of human behavior during an emergency evacuation, we have shown that emergency guidance robots can have a significant positive effect on survival rates. In this experiment, we assumed that information would easily flow through a crowd. We also neglected the possibilities of leaders who would sway the opinions of their group. In the next section, we present a model of information propagation that accounts for the flow of information in a crowd and simulates the effect that leaders can have on other evacuees.

3.3 Information Propagation in Emergencies

A fire in a crowded club is a frightening and confusing situation. Where is the nearest exit? Should the man shouting directions be trusted? Who is believable? Over the years many such fires have happened in crowded clubs and bars. For example, in The Station Nightclub fire of 2003, emergency personnel arrived within five minutes of onset of the fire, yet were helpless to prevent one hundred deaths [32].

For evacuation guidance robots to be effective, they must be trustworthy. Rushed and possibly panicked evacuees will not follow directions from a source they do not trust. In The Station Nightclub fire, evacuees followed directions from nightclub employees, policemen, and firemen because they trusted those sources [32].

In the previous section we evaluated the use of robots as guides in similar emergencies, but before considering robot guidance, we must first understand how people communicate necessary information among each other in an emergency. Information about viable exits must propagate in some way during an emergency, whether it be directly through verbal communication or indirectly through gestures and movement. In any reasonably sized group, there will be individuals who will be thoroughly convinced that their memory of the best exit is correct and thus will not be receptive to the opinions of surrounding evacuees. We begin by determining what ratio of such true believers to uncertain individuals were present during The Station Nightclub fire. Then we add guidance robots to the scenario and vary the percentage of evacuees who believe a robot's instructions to find the minimum percentage necessary for significantly better survival rates.

3.3.1 Background Information

Various standards organizations have performed extensive studies after mass casualty evacuation events. One such study was performed by NIST after a fire in The Station Nightclub killed 100 people [32]. This study decided that two of the main reasons for the high casualty rate was the fast spread of the fire and a major stampede at the main exit. The fire department was able to respond within five minutes of ignition, yet the fire was so bad that no firefighters could enter the building until the entire fire was extinguished. As much assistance as possible was rendered at the exit points and windows, yet a majority of the people who were able to escape still had injuries requiring hospitalization and few who escaped more than one minute after the start of the evacuation survived. The NIST simulations were later corroborated and extended in [19].

Recently, research has examined the influence that committed minorities (or true believers) has on a larger population of people [81]. The researchers found that just 10% of committed minorities can sway the entire population and consensus among participants occurs much faster with committed minorities. They further hypothesize that their results may explain the committed minority phenomenon that sociologists have noted elsewhere in politics and culture. This work follows other investigations into social consensus and alignment [29, 18]. We use this work as a starting point for developing our model of information propagation in evacuations.

3.3.2 Methodology

In order to find the critical percentage of evacuees who need to believe information from an evacuation robot, we first defined a model for human movement during an evacuation. Next, we defined a

model for how information about exit locations propagates during the evacuation. True believers were created to act as people with unswayable beliefs in an exit location. At this point, the model was tested to determine what ratio of true believers existed in The Station Nightclub fire. Then, a robot policy was created specifically for our fire simulation. Finally, the information propagation model was modified to allow robots to give exit beliefs to evacuees. The ratio of humans willing to believe information given by the robot was varied to determine the effect on survival rates.

3.3.2.1 Simulation Environment

All experiments took place in a simulation of The Station Nightclub fire (the simulation environment can be seen in Figure 3.7, real nightclub design in Figure 3.6). The nightclub is simulated as the combined area of the three large rooms: the main room, the sun-room and the bar. No interior furniture or stages were included. The hallway that caused many of the casualties is simulated as coming out of the front of the nightclub instead of contained within because the model of human motion used in this experiment was not robust enough to handle interior walls in an efficient manner. The main room is 16.6 meters by 10.9 meters and connected to the sun-room by 10.9 meters of open space [32]. The sun-room is 10.9 meters by 4.6 meters. The bar is 7.6 meters by 8.5 meters and is connected with the other rooms by a large passage. The main exit is between the sun-room and the bar. Two emergency exits accessible to patrons were simulated, one from the bar and one from the west side of the main room.

3.3.2.2 Human Behavior

According to [7], each person in an evacuation has an exit in mind at the start of an emergency. Most, if not all, people will use the front doors as a first choice. This was modeled by giving each person a belief that there was an exit (e_0) at a two-dimensional Gaussian random perturbation ($g(\mu, \sigma)$ below) from the main entrance (Equation 3.1). If an individual happened to see an exit along the way, they transferred their exit location belief to that exit. Some individuals are "true believers" who cannot be easily convinced that their exit is wrong. Other individuals are unsure of their chosen exit such that they are easily swayed by true believers and those who follow the true believers. This was modeled using a confidence parameter (c_0) where true believers and those who can directly see exits had 100% confidence and others had 0% confidence initially (Equation 3.2).

$$e_0 = e_{main} + g(\mu = 0, \sigma = 4.2). \tag{3.1}$$

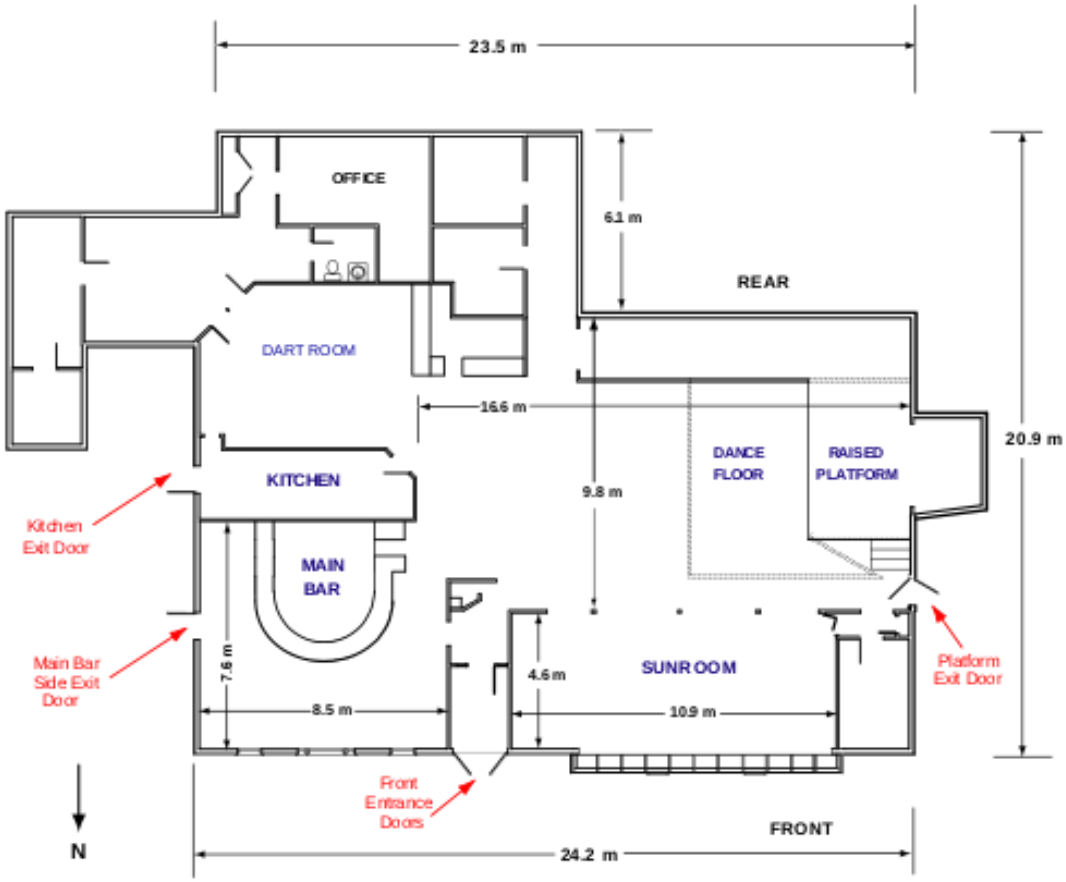


Figure 3.6: Actual Station Nightclub floor plan

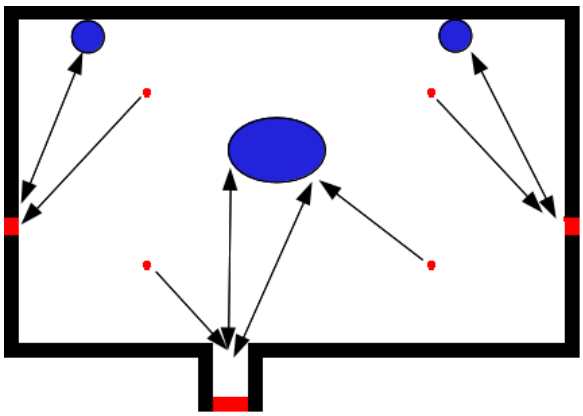


Figure 3.7: Directional information given to humans by robots. Exits are represented as red rectangles, robots as red squares, and holding areas as blue ovals. Directions given by robots are represented as arrows.

$$c_0 = \begin{cases} 1.0 & if\, true\, believer \\ \\ 0.0 & otherwise \end{cases}. \tag{3.2}$$

Each individual finds the unit vector ($\hat{v}$) from their current position (x_i) to their chosen exit (e_i) at each iteration of the simulation (Equation 3.3). The individual then attempts to move along that vector at his or her particular speed (s). If this movement causes collision with an obstacle or another person then the individual perturbs his end goal using Gaussian noise with a mean of 0 and a standard deviation of 1 distance unit (approximately $\frac{1}{12}$ of a meter) and a step along that path is again taken. If this perturbation still fails to place the individual in an open area then another perturbation is applied to the original goal. If ten tries fail to produce an open space then the individual is considered to be blocked in his or her original position for this iteration (Equation 3.4).

$$v = x_i - e_i. \tag{3.3}$$

$$x_{i+1} = \begin{cases} x_i + s * \hat{v} & if\, clear \\ x_i + s * (\hat{v} + g(\mu = 0.0, \sigma = 0.08)) & if\, blocked \\ x_i & otherwise \end{cases}. \tag{3.4}$$

3.3.2.3 Information Propagation

To model the propagation of exit knowledge through evacuees, we assume that individuals are capable of communicating the information they have about their exit as well as their confidence in their memory. In a real emergency it is unlikely that every person actually tells each other person where their exit is, however some information is exchanged by observing the trajectory of another person and any facial expressions he or she may be exhibiting.

Each person has a neighborhood that sets a range limit on how far information can be exchanged. Within this neighborhood, each person compares his or her exit with every other person's exit. The maximum confidence (c_y) is chosen as best, according to Equation 3.5. If this maximum confidence is less than the individual's confidence (c_i) then the individual will take the new exit (e) (Equation 3.6). Each time a person accepts a new exit location he or she also accepts the confidence degraded by a factor (a) (Equation 3.7).

$$c_y = max_{x \in N}(c_x). \tag{3.5}$$

$$e_{i+1} = \begin{cases} e_y & if\ c_y > c_i \\ e_i & otherwise \end{cases}. \tag{3.6}$$

$$c_{i+1} = \begin{cases} a * c_y & if\ c_y > c_i \\ c_i & otherwise \end{cases}. \tag{3.7}$$

3.3.2.4 Robot Behavior

This experiment was not designed to test different robot behaviors, so the behavior of the robots was kept as simple as possible. Each robot was given two locations to oscillate between. Each location had a corresponding direction that the robots gave the humans. Directions were given with the goal of keeping people on course to the closest exit while also keeping congestion low at the exit itself. This was achieved by having the robot alternate between directing people to a holding area and directing people to a given exit. Holding areas were chosen manually before the simulation. For this environment, three holding areas were chosen: one in the center, one on the left side of the top wall and one on the right side of the top wall. These areas were found to be sufficiently spread out to keep the humans in three groups but sufficiently close to allow individuals to move from the holding areas to the exit quickly. A holding area was chosen for each exit. Two robots were assigned to the main entrance and its holding area because the entrance is twice as large as other exits and thus could handle additional people guided by a second robot. The initial robot positions as well as the directional information pointing to the holding areas and exits are all shown in Figure 3.7. The robots were assumed to be simple platforms that could not detect humans as anything but obstacles and simply worked on timers to change places. Robot obstacle avoidance was implemented in the same way as the simulated human obstacle avoidance.

3.3.2.5 Human-Robot Information Propagation

For this experiment, we assumed that some humans will believe the robots, but others will ignore or even consciously disobey. We have modified the human exit information propagation model to determine how many evacuees must believe the robots to produce a significant change in survival rate. The humans who do believe the robots ($c_r = 1.0$) are modeled such that they become true

believers in whatever direction the nearest robot is advising. Their exit (e_{i+1}) is set to whatever direction the robot is giving (e_r, Equation 3.8). Their confidence (c_{i+1}) is set to maximum and they propagate information to other humans as before (Equation 3.9). In the case where an individual is a true believer and a robot believer, the robot's directions take precedence.

$$e_{i+1} = \begin{cases} e_r & if\ c_r = 1.0 \\ e_i & otherwise \end{cases}. \tag{3.8}$$

$$c_{i+1} = \begin{cases} 1.0 & if\ c_r = 1.0 \\ c_i & otherwise \end{cases}. \tag{3.9}$$

3.3.3 Human to Human Belief Propagation

Before an experiment can be run using robots, we must first determine a valid number of true believers for The Station Nightclub fire. A simulation of the nightclub and 440 people was created.

3.3.3.1 Experimental Setup

Simulations were initialized with 440 people inside the club. Each person was given the ability to see exits, robots and humans at a range of 3.5 meters. Each person's neighborhood was defined as all other individuals in sight. The confidence degradation constant was set to 0.9. Initial true believers were varied from 0% to 100% of the population at increments of 10%. Positions at the start of each experiment were randomized. The random seeds were kept such that each experimental setup could be run at each variable level. Thirty trials were run for each independent variable combination. The average human walking speed is approximately 1.4 m/s, so each human was given a speed within a Gaussian random position about that mean. It is estimated that all survivors evacuated the nightclub within one minute of the fire alarm, so experiments were run until one minute of simulated had time passed. Each iteration of the simulation was $\frac{1}{16}$ of a second in simulated time. The measured results of each test was the number of people who successfully evacuated within 1 minute.

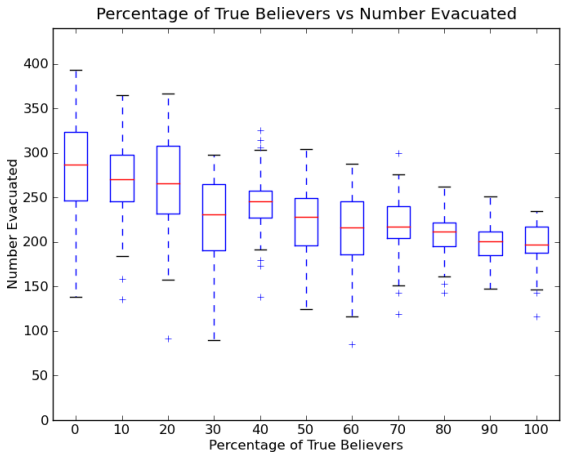


Figure 3.8: Results of human to human tests

Table 3.1: T-Test results comparing human to human tests and actual survival

Percent True Believers	P-Value
0%	<0.001
10%	<0.001
20%	<0.001
30%	0.418
40%	0.009
50%	0.506
60%	0.245
70%	0.767
80%	0.013
90%	<0.001
100%	<0.001

3.3.3.2 Results

The results of the human to human belief propagation tests can be seen in Figure 3.8. As the percentage of true believers goes up, the number who evacuate goes down. In other words, when more people listen to the others in their area instead of their own intuition, the survival rate goes up.

In the actual Station Nightclub fire approximately 220 people were able to escape through the doors (the others left through broken windows). Only door evacuation was simulated here, so a t-test was performed (Table 3.1) to see which percentage of true believers had $p > 0.05$ when compared with the actual number of survivors. This tells us that the total number of true believers in the nightclub was likely between 30% and 70%. This is a large range; however, the data taken from the actual event has a large error margin which prevents limiting the range.

3.3.4 Robot to Human Belief Propagation

Evacuation robots were added to the simulations with true believer rates not statistically significant when compared to the actual fire (30% to 70%) to determine what effect the robots had on survival rates. Tests were also run at 0% and 100% of true believers to determine the effects of robots on extreme populations.

3.3.4.1 Experimental Setup

For these simulations, humans were randomly chosen to either believe the robots or not. The percentage of humans who believed the robots was varied between 0% and 100% at 10% increments for each chosen true believer level.

Each robot was given a set of waypoints to move along to inform as many people as possible. Each robot was also given directional information to give to each human in range. The directional information for each robot is shown in Figure 3.7. Two robots were assigned to guide people towards the front entrance and one was assigned to each side entrance. For this simple robot model, the directions given were static and time based. The information was selected such that the evacuation of this particular nightclub would be optimized. It is assumed that any system that implements this work in the real world would take the time to customize directions based on their particular evacuation plan.

3.3.4.2 Results

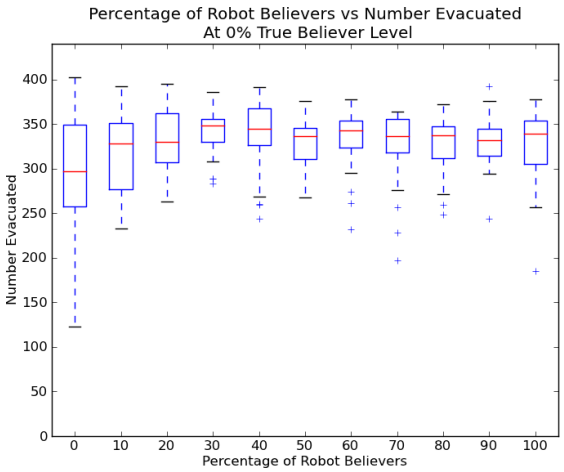
With no true believers, any number of humans believing the robots significantly increased the survival rate (Figure 3.9a). Table 3.2 shows the p-values as compared to the without robot trial. Every trial with any humans believing robots was significant at the 0.05 level. Robot believer ratios between 20% and 90% were significant at the 0.001 level.

The results of the 30% true believer test can be seen in Figure 3.9b. Here the robots had a significant impact on survival rate at the 0.001 level for all robot believer rates 10%-90%.

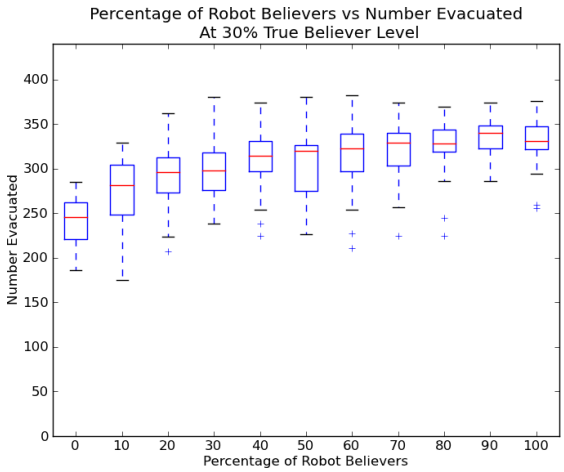
The results of the 40% true believer test can be seen in Figure 3.9c. The results become significant when 30% of the people believe the robots. Results at 40% true believers are very similar to those at 50-70% true believers, so those graphs have been omitted for brevity.

At the 100% true believer level (Figure 3.9d), results are significant starting at 10% robot belief ratio.

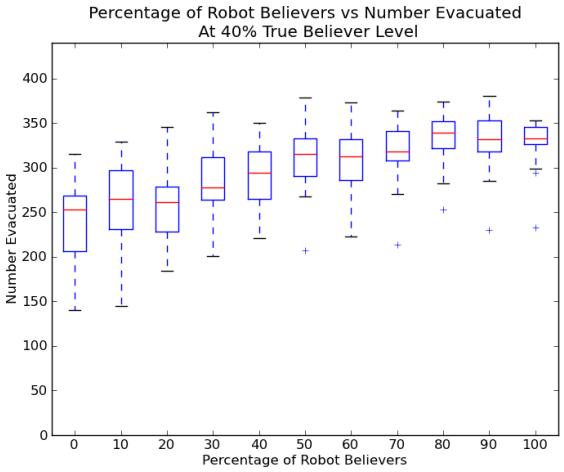
Table 3.2 shows the p-values from the t-tests between each of the conditions at each value with



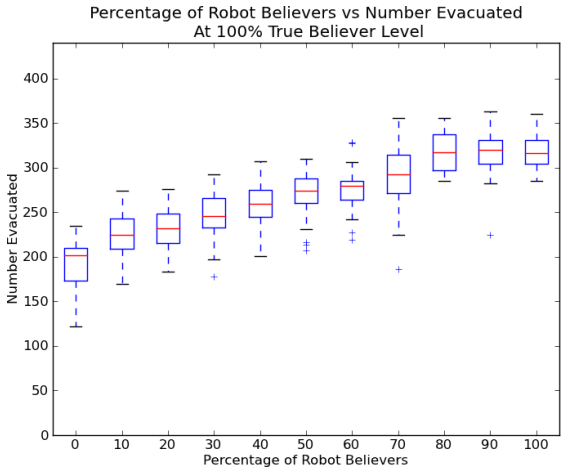
(a) Results of robot to human tests at 0% true believer level



(b) Results of robot to human tests at 30% true believer level



(c) Results of robot to human tests at 40% true believer level



(d) Results of robot to human tests at 100% true believer level

Figure 3.9: Results of selected exemplars of robot to human tests

Table 3.2: T-Test results of robot to human belief propagation tests compared with non-robot tests

		True Believer						
		0%	30%	40%	50%	60%	70%	100%
Robot Belief	0%	0.291	0.114	0.953	0.662	0.352	0.682	0.458
	10%	0.007	<0.001	0.107	0.010	0.004	0.360	<0.001
	20%	<0.001	<0.001	0.118	<0.001	0.001	<0.001	<0.001
	30%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	40%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	50%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	60%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	70%	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	80%	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	90%	0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	100%	0.005	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

the corresponding non-robot test.

3.3.5 Discussion

As a lower bound, just 30% of the humans have to believe the robot to increase survivability at a statistically significant level. As more people believe, we see a dramatic rise in survivability. When all humans believe the robots, over 100 extra people can make it out within the time limit. If we extrapolate to include window evacuations during the actual event then it is possible that all people would have made it out of The Station Nightclub in the 2003 fire if robots were available to help. In general, standard deviation also dropped when robots were introduced, so more people consistently made it out.

In extreme cases where either no humans are true believers or where all humans are true believers, robots have a significant impact on evacuation rates at the lowest level of robot believers tested. Based on these results, it may be true that evacuation robots will be most helpful in areas where most people believe they know the best exit, such as an office building. The people would almost all head to the front entrance, but the robots could guide some to side exits. Likewise, in areas where the location of the nearest exit is unknown, such as large malls and convention centers, the robots can provide much needed guidance.

Adding robots to an evacuation introduces the risk that the robots themselves act as obstacles. For all percentages of true believers, holding robot belief at 0% produced no significant results when compared to tests when robots are not present. From this, we can conclude that the robots' presence had no effect on the simulation unless the people believed in the robots, thus the robots did not produce a noticeable impediment in the evacuation when they were ignored.

3.4 Conclusion

Our experiments indicate that only 30% of a group needs to be convinced to follow the robot in order to save the whole group. By examining group affinity in evacuations we know that robots must be aware of family groups that will refuse to evacuate without the entire group. Additionally, the first experiment shows that robots can produce a significant positive effect in an emergency. By examining after-disaster reports we know that a large percentage of casualties were unable to find an exit in time or tried to exit through an already congested area. As such, emergency guidance robots need to communicate understandable guidance instructions in a trustworthy manner to be useful in evacuations. In the next chapter, we develop prototype robots for this task.

Chapter 4

Prototype Emergency Guidance Robots

4.1 Introduction

In a fire emergency, evacuations are usually triggered by alarms. These alarms have an audible component (usually a horn or a voice) and a strobe light. The functional purpose of the system is to provide notification and guidance to facilitate an evacuation in a fire emergency. The goal of the system is to minimize response time as well as total evacuation time. In current emergency scenarios, guidance is typically provided visually through exit signs, but notification is provided visually and audibly using fire alarms. According to the Boyce study [12], exit signs have a maximum visibility of approximately fifteen meters and, according to the NIST review [21], this is greatly reduced as smoke fills the room. Exit signs and fire alarms are supplemented by published evacuation plans. Plans are typically posted in places where people unfamiliar with the building will see them, like hotel rooms and conference venues. While evacuation plans are usually publicly available and posted in prominent locations, visitors may not be able to locate them in an emergency and are unlikely to have studied them prior to an alarm.

In the previous chapter, we established that guidance robots can have a significant positive effect on emergency evacuations if they are understood and trusted by the evacuees. Designing a robot that can be understood in such a chaotic environment is not a trivial task. The design must consider that evacuees will be interacting with the robot in an environment that is potentially smokey, crowded, noisy, or all three. The design must also appear to be trustworthy to evacuees. Based on a review

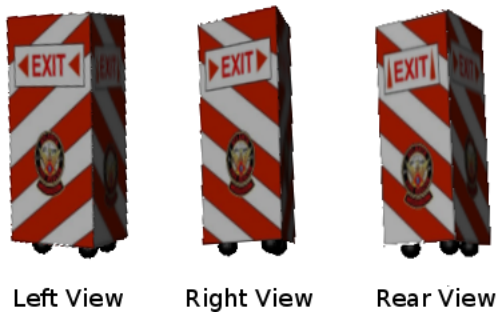


Figure 4.1: Emergency Guidance Robot Prototype 1

of existing emergency guidance technology, two prototype robot designs were created. The designs included a number of the recommendations from the literature including red and white colors, clear arrows, and lighted signs. These designs were then evaluated with participants in a virtual simulation of a fire emergency.

4.2 Design

The first robot was designed with three sides (Figure 4.1). The rear side was designed to be noticeably narrower than the other two so that the robot's forward direction was clear. The three sides of the robot were nearly identical, except for directional arrows. The robot was at least as tall as an adult human so that it could be seen in a crowd. Each of the top three corners had a downward facing light to illuminate the area around the robot. These were very bright to help evacuees see where the robot was and where they were going. Flashing lights and strobes were avoided as the evacuees were expected to look at the robot as they followed it to an exit.

The most important aspect of the design was the static display featured prominently towards the top of each side of the robot. This is shown in the diagram as a standard North American exit sign. The color and style of the sign would be changed based on the location of the robot. In Europe, this would be a green sign with a figure heading towards the front of the robot. All signs, regardless of style, had directional arrows that pointed towards the front of the robot. It was assumed that the front of the robot would always be pointed towards the best exit path. The exit signs were illuminated brightly from behind, but not so brightly as to blind people within a meter of the robot.

To encourage trust, the robot was designed with red stripes to make it resemble a fire truck. The robot would need approval from national and local fire safety organizations, so their logos were shown as large as possible to convince evacuees that the robot was an approved evacuation device. If possible, the local fire department would put their logo on it as a seal of approval.

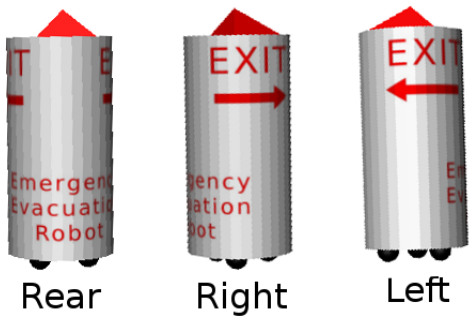


Figure 4.2: Emergency Guidance Robot Prototype 2

A second robot model was created after receiving reviews from other researchers (Figure 4.2). The changes in this robot were motivated by concerns that the first robot looked like a static sign when it was still, rather than a dynamic robot. The same white and red colors were used, although the stripes were not included in this model. The robot had "EXIT" written twice on either side of its cylindrical body with arrows pointed towards the front. There was a three dimensional arrow on top also pointed towards the front of the robot. This arrow was in response to comments that the first robot model's forward direction was somewhat ambiguous. "Emergency Evacuation Robot" was written along the back to make the robot's purpose obvious.

4.3 Evaluation

To evaluate these prototypes, we created a three dimensional environment using the jMonkeyEngine 3 game engine to simulate an emergency and determine to what degree an individual will follow a robot to a variety of exits. This engine was chosen so that the simulator could be deployed in a web browser as a Java Applet. A small shopping mall environment and two robot models were created in Blender. Test subjects started near the front entrance and were instructed to proceed to a highlighted region towards the back of the mall in their own time. Once the participant entered this region, smoke filled the mall and the interface displayed the text "FIRE EMERGENCY! EVACUATE!" This gave the participants time to explore the environment before the emergency, as if they were in a real mall. The mall was left empty of people and obstacles to better study the effects of one individual in an emergency. All exits were marked with exit signs in front of the exit and in any hallway leading to the exit, as in a normal mall.

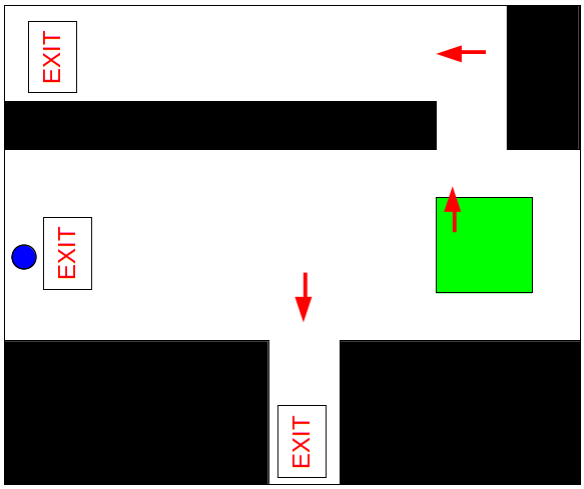


Figure 4.3: Map of the mall environment. Exit signs indicate the location of exits. Red arrows indicate directional exit signs (and the direction they point). The blue circle indicates the starting position of the participant. The green square indicates the highlighted region.

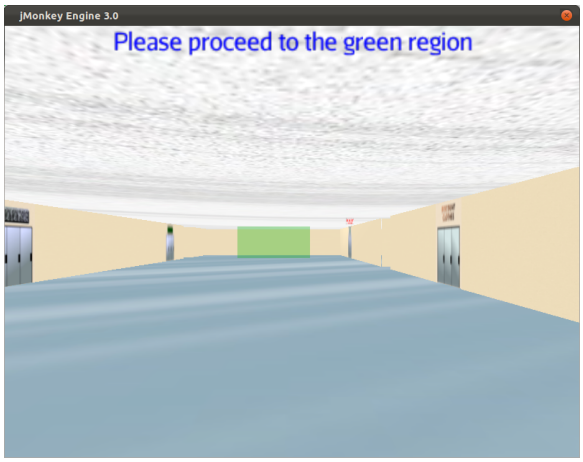


Figure 4.4: Participant view at start of simulation. Highlighted region is immediately ahead.

4.3.1 Environment Model

A small shopping mall environment was created to test the prototypes (Figure 4.3). Three exits were created to give the participant a choice during simulations. The exits were each approximately the same distance from the highlighted region. Each exit was marked with exit signs at each corner. The front exit was just behind the starting position of the participant (Figure 4.4). This gave the appearance of the participant entering a mall through the main doors. Another exit was to the left approximately halfway down the main area of the mall. The final exit was along a corridor immediately to the right of the highlighted region. Without any additional guidance and with no smoke obstruction, the participant was expected to move straight towards the large exit in front.

Some attempt was made to give the appropriate atmosphere to the simulator by adding store-

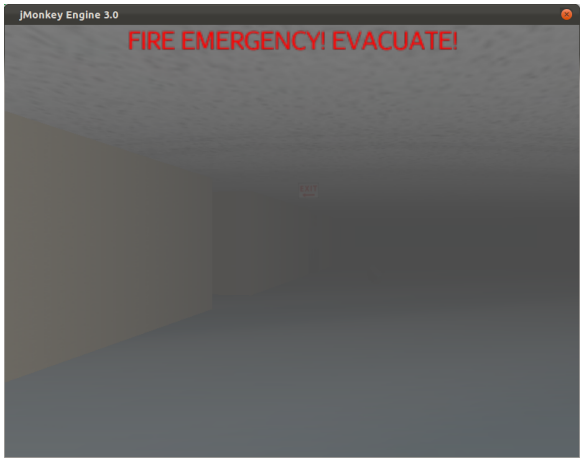


Figure 4.5: Participant view at start of emergency.

fronts and textures along the wall, floor and ceiling. A simple outdoor scene could be seen out of each door, along with some sunlight. A patterned texture was used on the floor and ceiling so that the participant would have a sense of motion as they moved through the scene.

Smoke was added to the model by using the game engine's fog mechanism with a dark gray color. The smoke level was set such that walls were just visible across the large hallway. A view of the simulation after smoke was added is in Figure 4.5.

Participants controlled the simulator with the arrow keys. Up and down were used to traverse, left and right were used to turn.

4.3.2 Robot Behaviors

Both robots followed the same control policy. The robots were assumed to have a holonomic drive such that they could point towards the desired exit regardless of their direction of travel. Robots were given a list of targets at which to point for each exit scenario. The targets were set at each corner along the desired path and at the final exit. The robot could choose between five positions in front of the participant's view: far left, middle left, center, middle right and far right. The robot chose whichever position was closest to the desired target and pointed towards that target. The robot attempted to stay within the social zone of the participant with a proportional velocity controller. The maximum speed of the robot was set at three times the maximum speed of the participant.

4.3.3 Hypotheses

Our first hypothesis was that evacuation times would be faster with a robot present than without. Previous simulation results have shown robots to be effective in an emergency and we specifically

Table 4.1: Scenarios

ID	Scenario
0-X	No robot appears
1-F	Robot 1 appears and instructs the participant to proceed to the front exit
1-L	Robot 1 appears and instructs the participant to proceed to the left exit
1-R	Robot 1 appears and instructs the participant to proceed to the right exit
2-F	Robot 2 appears and instructs the participant to proceed to the front exit
2-L	Robot 2 appears and instructs the participant to proceed to the left exit
2-R	Robot 2 appears and instructs the participant to proceed to the right exit

designed these robots and behaviors with that in mind (Chapter 3). Our next hypothesis was that Robot 1 would be better received than Robot 2 in the survey results. Our intuition was that Robot 1 presented a greater sense of urgency to the participant and thus was a better evacuation robot. We included Robot 2 in this experiment to test this intuition. Our final hypothesis was that most participants would follow the robot initially, but few (if any) participants would follow every time. We expected that the first time the robot appeared the participant would follow it out of curiosity. Some participants were likely to follow it several more times out of trust. Eventually, the robot would go in a direction that the participant did not think was safe, or the participant would become fatigued with the test and proceed on his or her own course.

4.3.4 Experiment Procedure

Each participant was asked to complete seven total scenarios, presented in random order (Table 4.1). Volunteers were solicited by sending an announcement with a link to the interactive simulations via email. Volunteers performed the test on personal computers outside of a lab environment.

Each scenario ended when the participant reached an exit. Participant and robot (when applicable) positions were recorded at 0.5 second intervals. The time at the start of the scenario, time at the start of the emergency and time when the participant reached an exit were all recorded for each scenario. After all of the scenarios, participants were given a short survey (Table 4.2). All Likert Scale questions were rated on a scale of 1 (labeled "Strongly Disagree") to 7 (labeled "Strongly Agree).

4.3.5 Results

4.3.5.1 Scenario Results

Fifteen volunteers completed all seven scenarios. Any volunteers who completed fewer than all scenarios were excluded from the results presented below. As can be seen in Figure 4.6, every

Table 4.2: Survey Questions

Label	Question	Response
Q1	In what year were you born?	Short answer
Q2	What is your gender?	Multiple choice
Q3	What is your occupation?	Short answer
Q4	I am comfortable with using new technology	Likert Scale
Q5	I believe firefighters are trustworthy guides in a fire emergency	Likert Scale
Q6a	Robot 1 [accompanied by picture] looked like a trustworthy guide.	Likert Scale
Q6b	How could this robot's appearance be improved to encourage evacuees to follow it?	Short Answer
Q7a	Robot 1 acted like a trustworthy guide.	Likert Scale
Q7b	How could this robot's behavior/motion be improved to encourage evacuees to follow it?	Short Answer
Q8a	I followed Robot 1.	Likert Scale
Q8b	Why or why not?	Short Answer
Q9a	Robot 2 [accompanied by picture] looked like a trustworthy guide.	Likert Scale
Q9b	How could this robot's appearance be improved to encourage evacuees to follow it?	Short Answer
Q10a	Robot 2 acted like a trustworthy guide.	Likert Scale
Q10b	How could this robot's behavior/motion be improved to encourage evacuees to follow it?	Short Answer
Q11a	I followed Robot 2.	Likert Scale
Q11b	Why or why not?	Short Answer
Q12a	During this simulation, I acted as if I were in a real emergency.	Likert Scale
Q12b	Why or why not?	Short Answer
Q13	What would make the simulation more realistic?	Short Answer
Q14	Is there anything else that would encourage you to follow a robot in an emergency?	Short Answer
Q15	Please list any other comments here	Short Answer

participant followed the robot at least twice. Five participants followed the robot in every scenario where a robot appeared. In general, participants followed the robot for the first few scenarios and then tended to leave by the left exit. This is most likely because they were frustrated from running many similar scenarios and perceived the left exit to be the closest from the highlighted region.

Three of the robot scenarios show a trend ($p < 0.1$ using a paired t-test) of allowing participants to evacuate faster when robots were present (see Figure 4.7 and Table 4.3). It should be noted that the averages include those participants who did not follow the robot, so we expect that a larger sample size would allow us to find significant results that show that following the robot produces a faster evacuation. There are no statistically significant results for scenarios between robots (Table 4.4), but the results for the front and right exits show a general trend where evacuation is faster with Robot 2.

4.3.5.2 Quantitative Survey Results

All volunteers reported being male students between the ages of 20 and 32 (mean of 24). Participants rated their level of comfort with technology at a mean of 6.1 and rated the trustworthiness of firefighters in an emergency at a mean of 6.9. Participants rated the two robots as essentially the

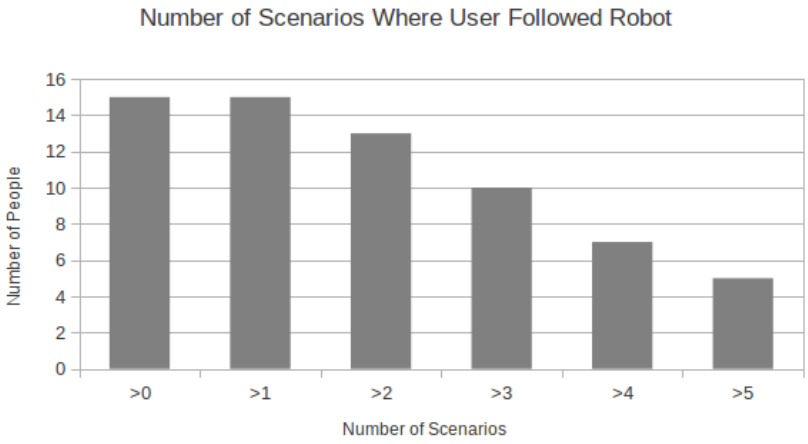


Figure 4.6: Number of scenarios where participants followed a robot

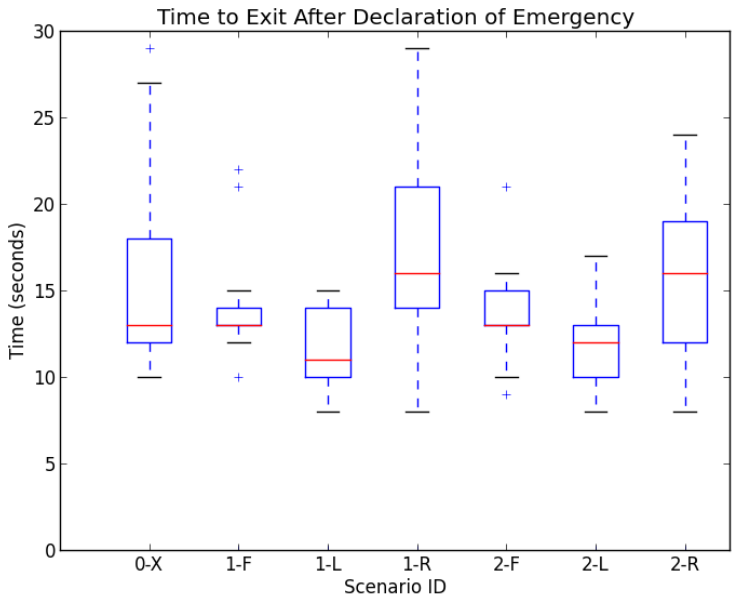


Figure 4.7: Average time from start of emergency to exit

Table 4.3: P-Values between no-robot scenario time to exit and other scenarios.

Robot	Scenario	P-Value
1	Front Exit	0.287
1	Left Exit	0.029
1	Right Exit	0.166
2	Front Exit	0.073
2	Left Exit	0.052
2	Right Exit	0.911

Table 4.4: P-Values between robot scenarios times to exit.

Scenario	P-Value
Front Exit	0.174
Left Exit	0.859
Right Exit	0.116

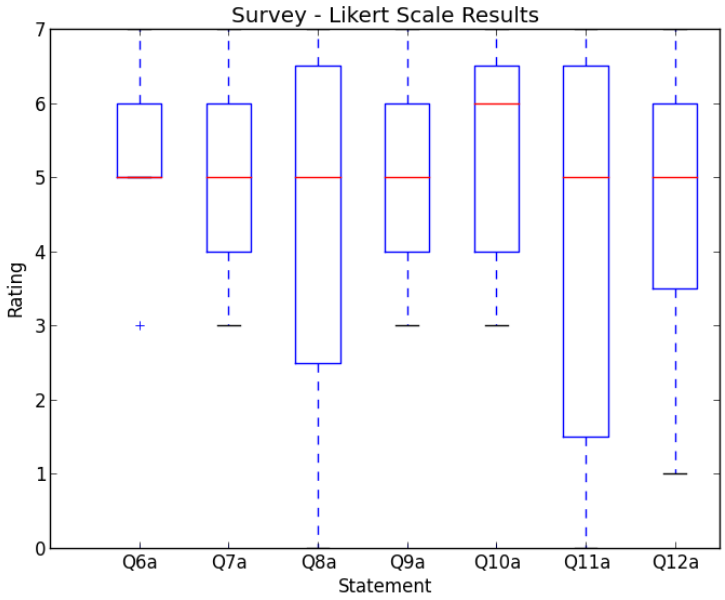


Figure 4.8: Survey Results

same on the Likert scale, with responses ranging from 5.1 to 5.6 on a seven point scale (Figure 4.8). Some participants failed to respond to some questions (Figure 4.5). It is unknown why the participants failed to respond. The realism of the simulation was rated with a mean of 4.7.

4.3.5.3 Qualitative Survey Results

Most of the free response comments given were constructive criticisms to help improve the design of the robots and simulation. The most consistent critique was that the robot would often move

Table 4.5: Number of Respondents Per Question

Question	Respondents
Q6	14
Q7	14
Q8	12
Q9	13
Q10	13
Q11	11
Q12	15

past a nearby exit to guide them to a further exit. Most participants realized this was by design as part of the experiment, but suggested that the robot should give some reason for not going to the nearest exit. Several participants suggested that including other evacuees would make the simulation more realistic. Five participants suggested that a more complicated environment would increase the necessity of a robot guide. Six participants said that they followed the robots until they understood the environment better, then took whichever way seemed fastest. One participant suggested that some sort of scoring system could be added to encourage a fast evacuation.

Robot 1 was generally well received, but several comments were made for improving its appearance. One participant commented that it looked like a "candy cane" and that it could be improved by adding explicit information that the robot was serving an emergency purpose. Several participants suggested that lights or strobes should be included on the robot. Most participants mentioned that audio notifications would help to increase the urgency of the emergency and help give guidance to evacuees. Three participants specifically noted that the exit signs on the side of the robot encouraged them to follow it.

Despite its slightly higher score on the Likert statements, Robot 2 received more negative comments. Several of the comments about the addition of lights and audio were repeated. One participant felt that the robot resembled a trash can. One participant noted that the robot was easier to lose in the smoke, but suggested that adding some color to it would fix that. Several participants mentioned that it was hard to tell which direction was forward, but some noted the same about Robot 1.

Two participants mentioned differences in actions between the two robots despite their identical programming. One thought that Robot 2 was faster and thus more trustworthy. One thought that one of the robots (he did not specify which) was attempting to deceive him. Two participants were confused because the robot turned too often. Four participants noted that the robot lost their trust when it passed a clearly marked exit in favor of one further away.

4.3.6 Discussion

Many of the comments given by the volunteers were used to create the next revision of evacuation robot designs seen in the next chapter as well as the evacuation scenarios shown in Chapters 6 and 7.

4.3.6.1 Robot Design

There was virtually no difference in the quantitative reviews between the two robot designs. This was something of a surprise, but because both robots received favorable reviews we anticipate that they both have features necessary in an evacuation robot. Thus far, audio has been avoided in the simulation because there is no guarantee that participants will have the speakers turned up on their computer. This could be solved by holding participant testing on a lab computer, but the intent is to distribute the simulator to the public for future testing. In our later work, messages from the robot were added to the simulator in the form of speech bubbles.

4.3.6.2 Robot Actions

Several participants complained that the robot passed obvious exits; however, that was intentional to determine how the participant responds to robot guidance so it will not necessarily be changed in the future. In response to this, we have defined distinct decision points where robots should be placed to provide guidance in an emergency. Instead of directing an evacuee to the exit, the robot will move to an intersection and inform participants which hallway to take.

4.3.6.3 Scenario Revisions

The two largest complaints about the simulator were the repetitive scenarios and simplistic environment. In Chapter 6, we have pursued a between subjects design that involves each participant completing a single trial. As an added bonus, participants are less likely to learn a simulation environment in their first experience with it, so even simple environments should provide some challenge.

When participants ignored robot guidance they tended to exit through the left door rather than the front door. This does not agree with [7]; in this experiment, the effect was probably caused by acclimation to the environment, but Chapter 7 explores this concept further.

4.4 Conclusion

Our first hypothesis was correct in three of the six scenarios. In those scenarios, participants evacuated faster with a robot present than without. Our second hypothesis was incorrect: participants did not rate Robot 1 significantly better than Robot 2. Both robots received favorable reviews on the Likert scale questions, so elements from both robots were used in later studies. Surprisingly, one-third of all volunteers followed the robots whenever presented, so our expectations in hypothesis

three were exceeded. While some of these results are certainly due to the novelty effect of evacuation robots, we are hopeful that the general public will accept these robots in the event of an emergency.

In previous work (Chapter 3), we have found that a lower bound of 30% of evacuees and an upper bound of 80% of evacuees must trust guide robots in order to produce significantly better survivability in an emergency. While our current sample size is too small for solid conclusions, this experiment has found that all tested individuals are willing to follow the robot in at least two scenarios and 33% of those tested followed in all scenarios. Thus, these prototypes were somewhat successful in attracting followers, but not all of their instructions were clear. The next chapter presents the next iteration of robot designs that use visual information conveyance modalities to instruct participants in an emergency.

Chapter 5

Emergency Guidance Robots

5.1 Introduction

Data from Chapter 3 indicates that conveying guidance information to a small percentage of evacuees drastically improves survivability. In the last chapter, we presented results indicating that our prototype robots were initially trusted by humans in a simulated emergency situation but that the trust levels dropped considerably when participants in the experiments were unable to determine what the robot was instructing them to do. Building on those results, in this chapter we explore various visual guidance modalities deployed on mobile robot platforms and their effect on human understanding of guidance instructions. We also redesigned the prototypes with commercial-off-the-shelf (COTS) parts to be more effective. We then tested each new design in a virtual environment, followed by verification tests with remote presence and physical presence robots. This work is in pursuit of the second contribution:

> *Developed models for communicating directional information to humans in high-risk, time-critical situations and identified their correlation to various robot form factors.*

5.2 Robot to Human Information Conveyance Modalities

Three methods for conveying guidance information were identified: static signs, dynamic signs, and arm gestures. These methods were combined on a mobile robot base to form five different platforms capable of conveying guidance information and one baseline platform with no specialized information conveyance ability. These systems were tested by recording simulations of the six platforms performing each of four different guidance instructions at an instruction point near the evacuee and

a point further away from the evacuee. Human participants viewed the information conveyed by the robot then interpreted the instructions and rated the understandability of the information being conveyed.

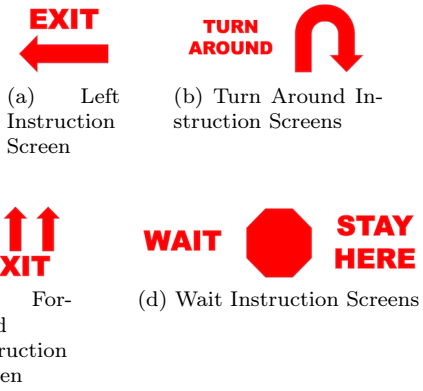
5.2.1 Modality Descriptions

This work is focused on methods to convey instructions to victims in a potentially noisy emergency situation. Some instructions in an emergency are directional: either instructing victims to go to a particular location or instructing them to stay in place. In this chapter, we only consider a victim standing in one location observing a robot giving instructions, so one set of instructions that we believe would be valuable is: 1) proceed to the left or right (we arbitrarily chose left in all cases), 2) proceed forward, 3) turn around, and 4) stay in place. The instructions as conveyed by each modality are given below.

Mobile Platform A typical mobile platform can convey information even when not equipped with specialized displays and actuators. Unlike in the previous chapter, we assume that the mobile platform is non-holonomic but otherwise a fully controllable ground robot. For directional instruction (left, forward, turn around) the robot first turns in the direction it wishes the human to proceed and then oscillates about that direction by 30 degrees left and right. In this way it can point in the general direction that the human should proceed but still indicate that information is being displayed through action. To instruct the human to stay in place, the robot spins in place.

Static Sign In the previous chapter, we mounted a static sign consisting of an arrow and the word "Exit" to a holonomic platform and pointed the robot in the direction of the exit. This was both confusing to the participants and unrealistic on actual platforms, so for this experiment the static sign only consists of information giving the intent of the robot by displaying the words "Emergency Guide Robot." The sign informs participants of the purpose of the robot but does not provide any specific instructions.

Dynamic Sign A dynamic sign gives the robot the ability to convey situation-dependent information including arrows, text, and animations. Such a sign can consist of a tablet or computer monitor or even a set of LEDs. For the purposes of this work, we assume that the sign will give sufficient resolution such that it can show English words and simple symbols. For the left and forward directions, the dynamic sign shows the word "EXIT" and an arrow or set of arrows that point in and



(a) Left Instruction Screen

(b) Turn Around Instruction Screens

(c) Forward Instruction Screen

(d) Wait Instruction Screens

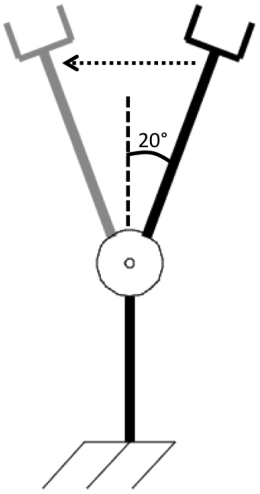
Figure 5.1: Dynamic Signs Text and Symbols

grow along the direction of the instruction (either left or forward, see Figures 5.1a and 5.1c). To indicate that the participant should turn around, the sign alternates between the u-turn symbol and the text "TURN AROUND" (Figure 5.1b). For the stay in place instruction the sign cycles through three screens: "WAIT," "STAY HERE," and a red octagon (Figure 5.1d).

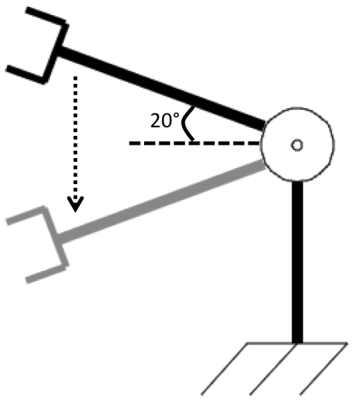
Arm Gestures Arm gestures are frequently used by humans in many different contexts, from police officers guiding cars to airport personnel directing aircraft to parents guiding children. Arms also provide the ability to attract attention at a distance by waving. We developed gestures for robots equipped with a single arm or multiple arms. For these purposes, we assume that the arms have at least two degrees of freedom: base rotation and at least one bend.

Attention is attracted by a platform with a single arm by holding it upright and waving it horizontally 20 degrees left and right (Figure 5.2a). For directional instructions, the whole platform turns to face the direction it wishes the human to proceed and the arm points forward (Figure 5.2b). The arm then oscillates slightly along the vertical axis to "wave" the participant in the required direction. For the stay in place instruction the robot faces the participant and waves its arm in the same manner it used to attract attention from this stationary position.

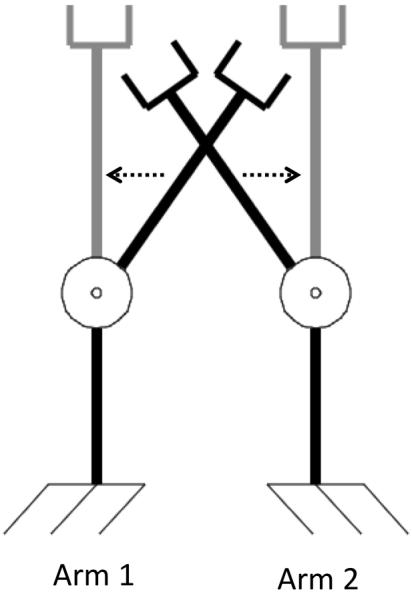
Multi-arm gestures are very similar to single arm gestures in directional instructions: two arms wave in the direction in which the participant should proceed. For the stay in place instruction, the robot faces the participant and alternates between both arms straight up and arms crossed (Figure 5.2c). Attention is attracted by waving upright arms.



(a) Example of a Single Arm Waving

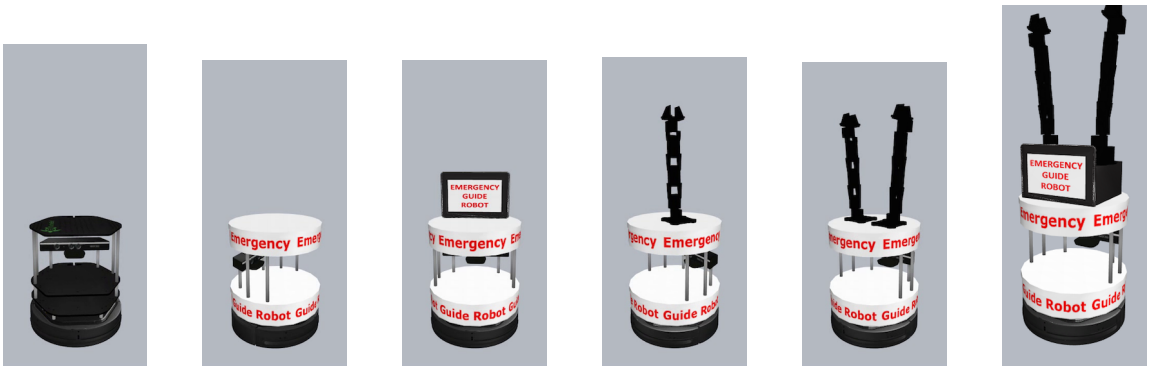


(b) Example of a Single Arm Pointing



(c) Example of Multiple Arms Crossing

Figure 5.2: Examples of Arm Gestures. In each case, the arm moves from the solid black position to the solid gray position in the direction of the dotted arrow.



(a) Baseline (b) Static Sign (c) Dynamic Sign (d) Single Arm Gesture (e) Multi-Arm Gesture (f) Multi-Arm Gesture with Dynamic Sign

Figure 5.3: Robot Guidance Platforms

5.2.2 Hypotheses

We hypothesize that a simple mobile platform will be unable to provide clear guidance instructions to humans but the addition of information conveyance devices, such as arms or displays, will allow for increasing understandability. The static sign is not expected to provide any specific guidance information but the dynamic sign is expected to produce significantly better results for the near case where text and symbols will be legible. For the far case, the arm gestures are expected to produce significant increases in clarity, with multi-arm gestures being more understandable than single arm gestures. Finally, we hypothesize that a combined approach featuring a dynamic sign and multi-arm gestures will convey guidance information best at both distance levels.

5.2.3 Robot Platforms

The platforms used in our experiment as well as the specific information conveyance devices are described in this section. We started with a Baseline robot (Figure 5.3a) to test the mobile platform instructions. We then created the Static Sign platform (Figure 5.3b) to determine if these signs produced any differences from the Baseline. All further platforms used static signs as well as other modalities. A Dynamic Sign platform (Figure 5.3c) as well as both Single Arm Gesture (Figure 5.3d) and Multi-Arm Gesture (Figure 5.3e) platforms were developed to test each of those modalities alone. Two arms were selected for the multi-arm platform to be as close as possible to human gestures. A final platform combined a dynamic display with multi-arm gestures to fully test our hypothesis (Figure 5.3f).

Mobile Platform All robot platforms were based off of the Willow Garage Turtlebot 2 due to its ease of use and general availability. The Turtlebot 2 is a 42 cm tall platform with a Kobuki base, a netbook running ROS for control and a Microsoft Kinect for sensing. The Turtlebot used in this experiment was simulated with 3D models of all components. This platform was tested without modification to determine the baseline understandability of guidance instructions.

Static Sign All robots except the Baseline carried signs that declared the robot's purpose as an emergency guidance aid. The signs were in two cylindrical components: one on the top of the Turtlebot and one covering the netbook just above the base. The top sign displayed "Emergency" in each of the four cardinal directions around the cylinder and the bottom sign displayed "Robot Guide" in the same manner.

Dynamic Sign An 11" Samsung Galaxy Tab was used as the dynamic sign. The tablet was mounted upright on top of the Turtlebot in landscape orientation. The tablet displayed instructions to the participant in a combination of arrows, stop-signs and English words.

Gesture Arms A PhantomX Pincher AX-12 arm was used in all platforms that required arms. This arm has five degrees of freedom and a maximum reach of 35 cm. For the Single Arm Gesture platform, the arm was mounted to the center of the top of the Turtlebot. For the Multi-Arm Gesture platform, two arms were used, one mounted on the left side of the top of the Turtlebot and the other on the right side. For the Multi-Arm Gesture with Dynamic Sign platform, the arms were mounted as in the Multi-Arm Gesture platform but on a box approximately 12 cm high such that no arm gesture would collide with the display.

5.2.4 Experimental Setup

To evaluate human understanding of the robot guidance modalities, we utilized a between-subjects experiment. Participants were recruited and the study conducted using Amazon's Mechanical Turk service. Other studies have found that Mechanical Turk provides a more diverse participant base than traditional human studies performed with university students [53, 15, 8, 37]. These studies found that the Mechanical Turk user base is generally younger in age but otherwise demographically similar to the general population of the United States (at the time of those studies, Mechanical Turk was only available in USA). A total of 192 participants performed this survey. Demographics of the participants can be seen in Table 5.1.

Table 5.1: Demographics of Participants (participants who did not answer specific questions are omitted from this table)

Category	Number of Participants
Total	192
Male	122
Female	69
Age 18-19	8
Age 20-29	114
Age 30-39	43
Age 40-49	13
Age 50-59	5
Age 60-69	4
Age 70+	1
Technical Profession	43
Customer Service Profession	16
Self-Employed	15
Unemployed	15
Clerical Profession	14
Other Profession	89
High School Diploma	22
Some College	60
Associate Degree	17
Bachelor Degree	68
Master's Degree	22
Professional Degree	3

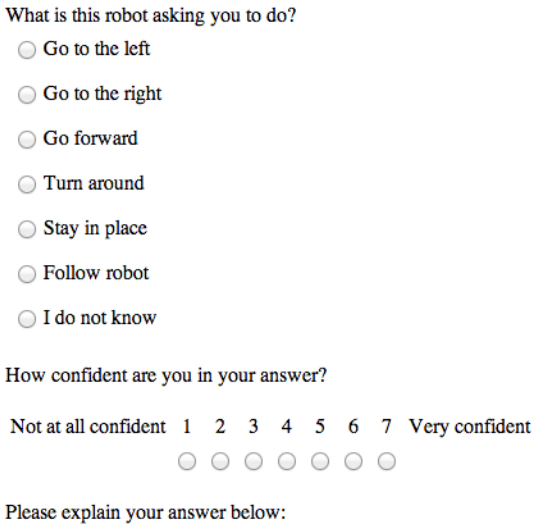
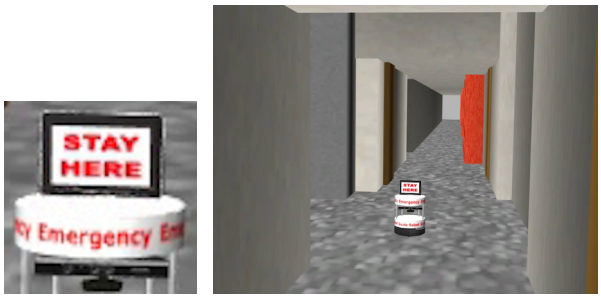


Figure 5.4: Questions asked for each video

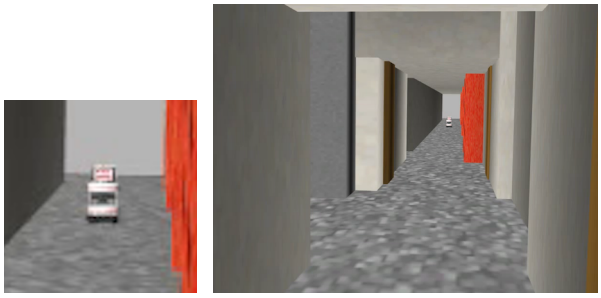
Participants began the study by reading and acknowledging a consent form. Next, they completed a demographic survey collecting information about gender, age, nationality (Mechanical Turk is currently available for residents of both USA and India), occupation, and education. Then, the participants were presented with videos of one particular robot performing each of the four instructions (one instruction for each video). A victim's ability to understand visual displays of guidance information depends on the distance between the victim and the display. For this reason, robots were tested at both a near and a far distance (see Figure 5.7 for the layout). Each participant was only shown the videos for one robot at one distance level. For each video, participants indicated which instruction they thought was being performed, estimated their confidence in that answer (a number 1 through 7), and gave an explanation for their answer (see Figure 5.4 for the exact questions and layout). Several instructions were given as multiple choice answers for each video, including some that never appeared in the test so that participants could not use process of elimination to give an answer. We recorded their answer to the multiple choice question and the explanation for choosing that answer. The order of the videos was randomized. The videos were each between 15 and 19 seconds long. Each video was 800 x 600 pixels in size. Participants were paid $0.50 for completing the survey. IRB approval was obtained before the study began.

Videos of the instructions were created in the Unity Game Engine. The videos were hosted on YouTube and embedded into the survey form on Mechanical Turk. Each commercially available component of each platform was simulated using CAD files provided by the manufacturers. These components were assembled into robot platforms using the Blender 3D modeling software and imported into Unity for simulation. Custom components, such as the signs on the robot, were created



(a) Actual size on screen (b) Entire scene

Figure 5.5: Dynamic Sign platform at near instruction point displaying wait instruction



(a) Actual size on screen (b) Entire scene

Figure 5.6: Dynamic Sign platform at far instruction point displaying wait instruction

in Blender and Unity.

The testing environment was a long hallway with open areas (potential exits) immediately to the left of the camera view, at the far end in front of the camera, and behind the camera. Screenshots of the robot in the near and far positions can be found in Figures 5.5 and 5.6, respectively. A map of the environment can be seen in Figure 5.7.

Sixteen participants viewed each robot at each distance level. To ensure that participants did not have any information about the other robots or their actions, no participant was allowed to perform the experiment more than once.

5.2.5 Results

In general, adding features to the baseline platform improved understandability of instruction (Figures 5.8 and 5.9). As expected, the Multi-Arm Gestures with Dynamic Display platform had the best overall understandability (75.8% overall) but, unexpectedly, the Static Sign platform performed worse than the Baseline (18.0% and 28.1%, respectively). Unfortunately, the confidence values re-

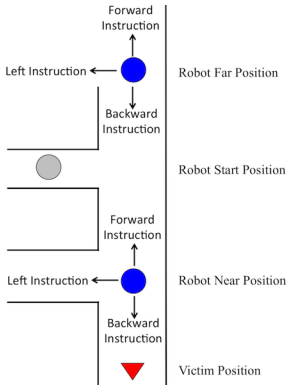


Figure 5.7: Map of testing environment

ported by the participants had no consistent base and thus could not be used to give insight into the results. Detailed results are given below.

We expected that there would be little or no difference between the Baseline and Static Sign platforms; however, the results show that the Baseline performed considerably better than the Static Sign for the left instruction at the near distance. Based on comments, it seems that participants were able to infer the rotation of the robot by the position of the Kinect. Because the top sign on the Static Sign platform partially obscured the Kinect, participants were not able to observe any orientation of the robot. Results from the other three instructions are very similar between the two robots. Participants were unable to see which direction the robot was pointing for the near distance forward and backward instructions, even when the Kinect was not obscured. For the far condition, participants in the surveys for both of these robots indicated that they could not see any understandable action. Many guessed that they should follow the robot (an option included in the survey even though it was not shown as an instruction). They reasoned that the robot moved away from them and waited, thus indicating that they should proceed in that direction. Overall, results from the Baseline and Static Sign platforms did not show statistical difference in a Chi-Squared test ($p = 0.054$). There were, however, significant differences between both of these robots' understandability and every other platform ($p < 0.001$) (Table 5.2).

As expected, the Dynamic Sign platform performed very well for the near instructions. Every participant indicated that the screen simply told them what to do and thus the answer was easy. The distance for the far instructions was specifically chosen such that a tablet or other screen about this size could not be clearly read from that instruction position. Each participant in the far survey confirmed that they were unable to discern any instruction from the Dynamic Sign platform, even when the large red octagon was displayed. Examples of the Dynamic Sign platform at near and far

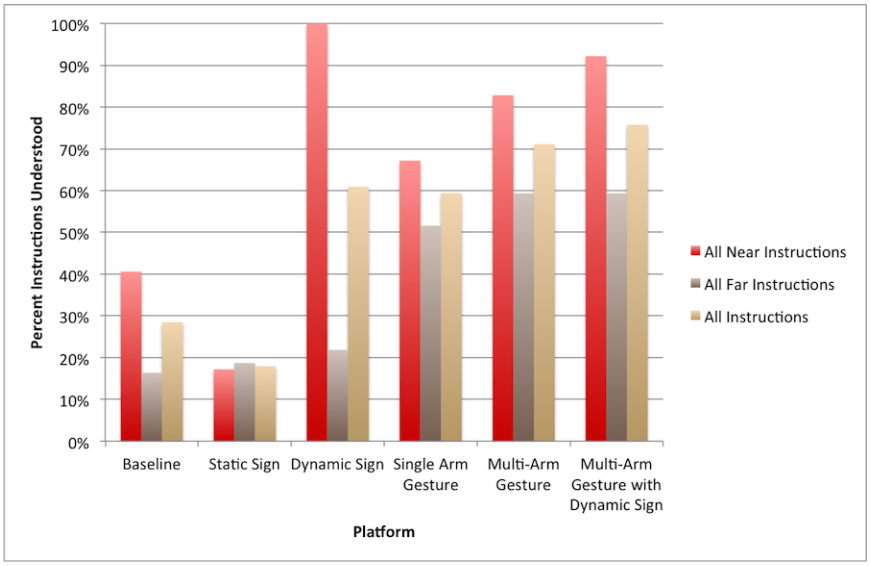


Figure 5.8: Percent Instructions Understood at Each Distance Level and Overall by Platform Type

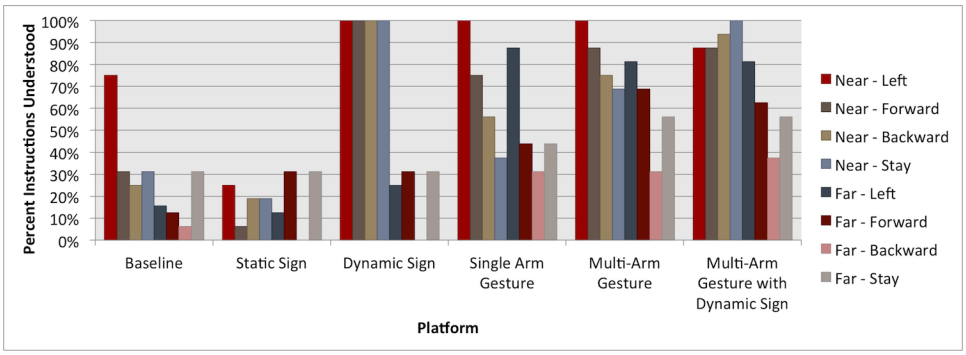


Figure 5.9: Percent of Instructions Understood by Platform Type

Table 5.2: Pairwise Chi-Squared Results Comparing Guidance Instruction Modalities (p-Values)

Platform	Baseline	Static Sign	Dynamic Sign	Single Arm Gesture	Multi-Arm Gesture
Static Sign	0.054				
Dynamic Sign	$<$ 0.001	$<$ 0.001			
Single Arm Gesture	$<$ 0.001	$<$ 0.001	0.798		
Multi-Arm Gesture	$<$ 0.001	$<$ 0.001	0.086	0.049	
Multi-Arm Gesture with Dynamic Sign	$<$ 0.001	$<$ 0.001	0.011	0.005	0.396

Note: $p < 0.05$ is significant

distances in Figures 5.5 and 5.6, respectively, show that participants in the far case were at a great disadvantage in this case. Recall that participants only viewed instructions from one robot at one distance level, so no participants were able to observe an instruction while close to the robot and then recognize it from a distance. A majority of the participants did not even realize that a display of any type was mounted on the platform. They wrote that there was simply an indecipherable red light on top. As in the previous cases, confused participants tended to assume that they should follow the robot if they could not understand a particular instruction. Overall, this platform was significantly different from the Baseline, Static Sign and Multi-Arm Gesture with Dynamic Sign (Table 5.2).

The Single Arm Gesture platform performed about as well at the Dynamic Sign platform overall (59.4% and 60.9%, respectively, $p = 0.798$) but had much lower variance between the near and far conditions. This indicates that guidance robots should be equipped with at least one arm unless the environment is conducive to individuals reading a screen. Participants had difficulty determining which direction the arm was pointing when it was giving forwards or backwards instructions. This difficulty increased with distance. Some of the difficulty could have been an artifact of the simulation. Participants also had difficulty understanding the stay instruction in both near and far cases. This is because the single arm is not able to articulate any standard stop gesture. In addition to the previously reported statistical results, the Single Arm Gesture platform had statistically significant

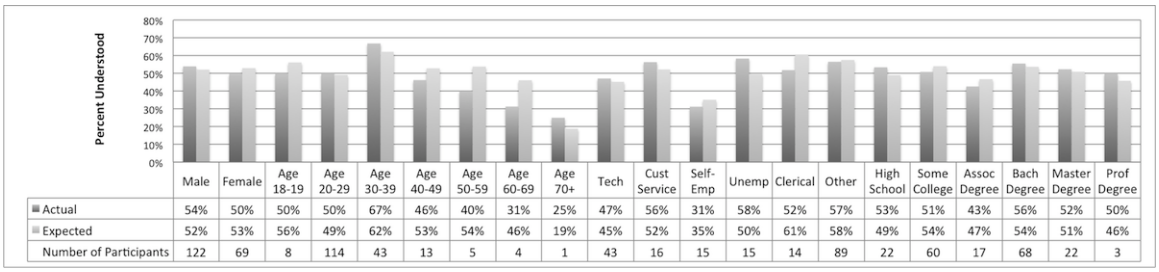
differences between the Multi-Arm Gesture ($p = 0.049$) and Multi-Arm Gesture with Dynamic Display ($p = 0.005$) platforms (Table 5.2).

The Multi-Arm Gesture platform solved the problem with forwards, backwards, and stay gestures by adding a second arm to provide instructions in the same style as airport ground crews. This produced an overall understandability of 71.1% of instructions. There was some confusion still with the forward and backward commands that was also exacerbated with distance, but comments indicated a greater confidence with the answers chosen. The stay instruction was confusing to some but most still recognized it as indicating to not proceed in that direction, even if they did not understand that they were supposed to stay in place. Otherwise, confusion generally resulted in the participant choosing the follow option.

Surprisingly, the Multi-Arm Gesture with Dynamic Display platform had no statistically significant differences from the Multi-Arm Gesture platform ($p = 0.396$). Overall, 75.8% of its instructions were recognized correctly. We expected the near results to be identical to the Dynamic Sign results, but comments from two participants lead us to believe they confused the robot's reference frame with the camera's reference frame and thought that the robot was indicating right instead of left and backward instead of forward. In those two cases, the robot also turns such that the tablet cannot be seen after the robot arrives at the instruction point, which might have increased the confusion. The robot performed as expected at the far distance level. Overall, after accounting for qualitative results gleaned from the comments, the combined approach of using a dynamic display and multi-arm gestures produced the best results for both near and far conditions.

Across all robot platforms the backward instruction was understood the worst (39.6%) and the left instruction was understood the best (65.6%). The instruction did have a significant effect on the results ($p < 0.001$) but all instructions were tested for each robot, so the results are still valid.

Recruiting participants through Mechanical Turk did not seem to have a major effect on the results. Most participants took the survey seriously and gave considered, thoughtful comments for each question. Some indicated frustration when they were unable to understand the robot. One even requested that participants receive training on robot gestures if we would like the results to be improved. There was some confusion as to exactly what the robot arms were, but the participants ability to understand the instructions did not depend on whether they referred to the arms as "antennas", "cranes" or even "tentacles." Only one participant gave bizarre answers, writing "I believe that the robot is trying to say that the walls are dirty and [that] they need to be cleaned." and "The reason I chose [to] follow [the] robot is because I think that the robot is attempting to communicate with the human." for two different questions. Those results were included for completeness.



	Male	Female	Age 18-19	Age 20-29	Age 30-39	Age 40-49	Age 50-59	Age 60-69	Age 70+	Tech	Cust Service	Self-Emp	Unemp	Clerical	Other	High School	Some College	Assoc Degree	Bach Degree	Master Degree	Prof Degree
Actual	54%	50%	50%	50%	67%	46%	40%	31%	25%	47%	56%	31%	58%	52%	57%	53%	51%	43%	56%	52%	50%
Expected	52%	53%	56%	49%	62%	53%	54%	46%	19%	45%	52%	35%	50%	61%	58%	49%	54%	47%	54%	51%	46%
Number of Participants	122	69	8	114	43	13	5	4	1	43	16	15	15	14	89	22	60	17	68	22	3

Figure 5.10: Results Grouped by Demographic Categories - Actual vs. Expected

A total of 122 males and 69 females participated in the experiment (one participant declined to give his or her gender). Gender was not found to have a significant effect on the results ($p = 0.183$). Participants spanned all education categories with a majority indicating that they had at least some college experience. This, too, was found to not have a significant effect on the results ($p = 0.758$). Most participants reported that they were in their 20s, but 10 were over 50 years of age, so the age range in this study is likely much more broad than would be found in testing on a college campus. Occupations spanned a wide range. We grouped them into the following categories for analysis: self-employed, technical, customer service, clerical, unemployed and other. Neither occupation nor age were found to have a significant effect on the results ($p = 0.441$ and $p = 0.446$, respectively). There was not enough variability in nationality to test for statistical differences. Expected results for Chi-Squared tests were calculated by taking an average of the results of all platforms weighted by the number of participants in that demographic who participated in that survey. See Figure 5.10 for all demographic results and their expected values.

5.2.5.1 Humanoid Guidance Modality

One of the strengths of the Multi-Arm Gesture robot is that the guidance methods created are applicable for robots with a simple mobile base (such as the one tested above) as well as for humanoid robots. To test this, we performed a follow-up study where we applied the gestures to a simulation of the Darwin OP robot and performed the same tests as above on Mechanical Turk (Figure 5.11). The Darwin OP has one fewer degree of freedom in its arms compared to the Pincher arms used above and has some size limitations that mean it cannot cross its arms over its head. To address this, the simulation of the Darwin was increased in size by 50% to be comparable to the Multi-Arm Gesture robot tested above and the stay in place instruction was modified to cross arms in front of the robot instead of over its head.

Our follow-up study was performed in the same manner as the previous study. Sixteen participants were asked to evaluate each of the four instructions at the near instruction point and

Figure 5.11: Humanoid Guidance Robot

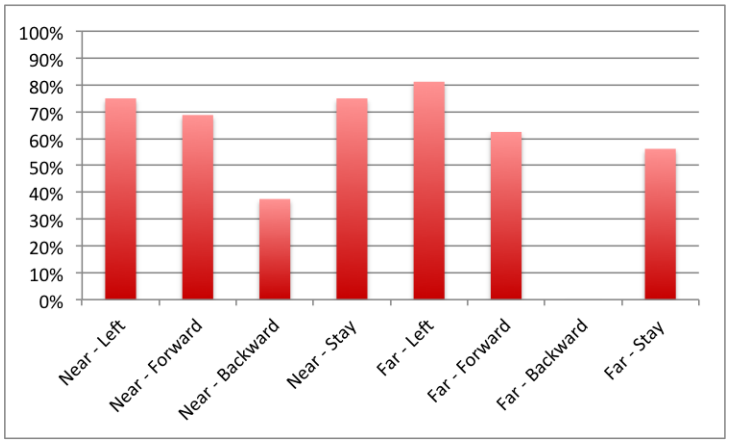


Figure 5.12: Percentage of participants who understood each direction at each distance level for humanoid guidance robot

another sixteen asked to evaluate the instructions at the far instruction point. This robot performed somewhat worse than the Multi-Arm Gesture robot above, even though participants generally understood the stay in place gesture (Figure 5.12). The directional instructions were more difficult to understand, possibly because of the smaller arms on the Darwin. The humanoid platform did not perform better than the other platforms, so we can conclude that a humanoid robot is not necessary to provide effective guidance.

5.2.6 Discussion

Our survey explored the capability of different robotic platforms to instruct humans to find a safe exit in an emergency. We focused on visual guidance to avoid potential problems with audio instructions in a noisy emergency environment. Platforms were varied by adding signs to indicate function, a tablet to display instructions in written language or recognizable symbols, and an arm or arms to gesture to the victim.

Through quantitative and qualitative results we found that a ground platform with a dynamic display and multi-arm gestures provides the clearest instructions to victims in an emergency. The

addition of single arm gestures or a dynamic display by itself also performed considerably better than an unmodified ground robot. A surprising result provides a word of caution to fellow roboticists: adding seemingly trivial aesthetics such as signs can produce differences in outcomes of human-robot interaction experiments.

This survey was a useful first experiment to evaluate our guidance modalities, but its generality is limited by the use of simulated, virtual robots in a virtual environment. In the next section, we present validation of a subset of these tests using physical robots.

5.3 Validating Information Conveyance Modalities with Physical Robots

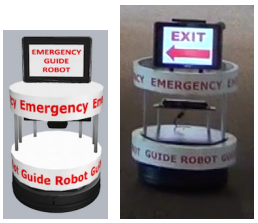
5.3.1 Introduction

Currently, human-robot interaction experiments are usually performed in a laboratory or real-world setting where robots can physically interact with human participants. Two alternatives to the traditional physical presence experiment paradigm now exist: a remote presence paradigm where the robot is located elsewhere so interaction occurs through video streaming and a virtual presence paradigm where the participant interacts with a simulation of a robot in a virtual environment. Some interactions, such as those involving emergency situations, are difficult to perform in a laboratory setting and can be impossible to perform in a real-world setting. Throughout this work, we have used virtual environments to evaluate prototypes of our emergency guidance robot. In this section, we determine the extent to which these virtual environments can be used to evaluate human understanding of instructions given by robots by comparing our experiment in the previous section with data from two new experiments. First, we used physical agents to create similar videos to those shown in our previous work, evaluated using crowdsourcing (remote presence experiment paradigm), and then we performed a similar experiment in a laboratory setting (physical presence experiment paradigm).

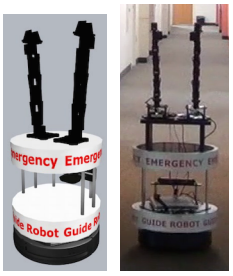
5.3.2 Experimental Setup

In this study, two additional experiments were performed and compared with the previous virtual experiment results. The first experiment tested the remote paradigm by recording videos of real robots performing gestures and using Mechanical Turk to gather data. The second experiment tested the physical paradigm by gathering participants in an office environment and measuring

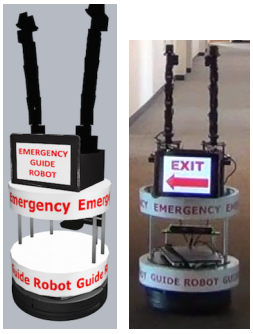
(a) Baseline Platform



(b) Dynamic Sign Platform



(c) Multi-Arm Gesture Platform



(d) Multi-Arm Gesture with Dynamic Sign Platform

Figure 5.13: Robots used in this study compared to their virtual counterparts. Virtual platforms are shown on the left and physical platforms on the right for each platform.

their response to a real robot performing instructions in front of them. Each experiment used the platforms or a subset of the platforms shown below (Figure 5.13).

5.3.2.1 Remote Presence Experiment

To evaluate human understanding of the remote robot guidance modalities, we utilized a between-subjects experiment. Participants were recruited and the study conducted using Amazon's Mechanical Turk service. A total of 128 participants performed this survey. These were compared to the 128 participants from the previous study who observed the virtual robot counterparts in the previous section. Participants answered the exact same survey questions as in the previous section. Again, two distance levels were used and the floor plan was the same as in Figure 5.7. Each participant only observed one robot at one distance level. The order of the videos was randomized. The videos were each between 38 and 89 seconds long. Each video was 1280 x 720 pixels in size. Participants were paid $1.00 for completing the survey. IRB approval was obtained before the study began.

5.3.2.2 Physical Presence Experiment

To evaluate our robot in a physical presence experiment we again used a between-subjects study. A total of 48 participants were recruited by posting fliers around the Georgia Tech campus and by emailing students in the School of Electrical and Computer Engineering. Only three conditions were tested in this study: the Baseline, Multi-Arm Gesture, and Multi-Arm Gesture with Dynamic Sign platforms were each tested at the near distance level. The near distance level (the robot was within approximately two meters of the participants) was chosen because our other experiments and robot use cases (Chapter 7) did not require participants to understand the robot at distances further than the near distance level.

The experiment began with participants reading and signing a consent form. Participants then lined up along a wall facing the robot's demonstration point in an office environment. The experiment location was in the same building as the videos for the remote experiment were recorded, but in a different location because the hallway used in the videos had considerable foot traffic that would disrupt the experiment. The wall on which participants lined up had several doors on it, so it was hoped that participants would consider one of those doors as a possible exit. A hallway in front gave the impression that participants could travel to the left, right, or forward to find an exit. Multiple participants observed the robot's demonstrations in each trial; however, participants were instructed not to communicate with each other during the procedure and an experimenter was present to supervise. Participants observed a single platform perform all four instructions and answered survey questions about each. The survey questions were identical to those in the virtual and remote experiments except on paper instead of a webpage. In our prior experiments we failed to find an ordering effect, so in this experiment we did not randomize the order of the instructions for each session. Each session observed first the backward, then the left, then the forwards, and finally the stay instruction. Participants were allowed to revise previous answers as long as the experiment was in progress. After the demonstrations, participants completed a demographic survey and were allowed to ask any questions they might have about the robot or the experiment. Participants were compensated $10.00 for their time. IRB approval was obtained before the study began. All robot demonstrations were automated. The experimenter controlled which demonstration would be presented at which time using a laptop interface.

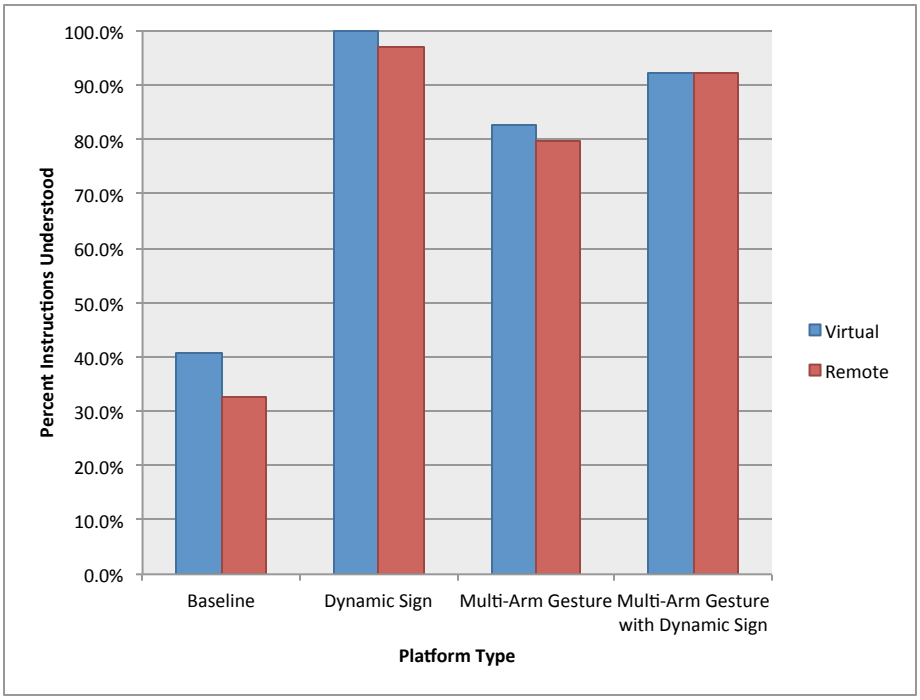


Figure 5.14: Results at the near distance level for the Remote Presence Experiment

5.3.3 Results

5.3.3.1 Remote Presence Experiment

A total of 128 participants (denoted as N below, 39% female, mean age of 34.1 years old) responded to a total of 512 questions (denoted as R below) in this study. Their answers were compared to answers ($N = 128, R = 512$) about the same platforms at the same distance levels in the virtual study performed. Results for each platform and comparisons with the corresponding platform are in Figures 5.14 and 5.15. Statistical tests were performed to determine any difference between the overall results across both distance levels and all instructions of a platform tested in the virtual experiment and the corresponding platform tested in the remote experiment.

Baseline Overall, 22.7% of instructions were correctly understood when presented by the remote Baseline platform. This is 5.8% worse than the virtual platform results, although the difference is not statistically significant ($\chi^2(1, N = 64, R = 256) = 0.592, p = 0.314$). As in the virtual study, the left instruction was generally understood but the other instructions were not. In the near case, participants interpreted the oscillating motion in the forward and backward instructions as the robot shaking its "head" to indicate "no." Many responses (13 of 32) indicated that participants interpreted this "no" to mean that they should stay in place and not proceed in any direction. A majority (10

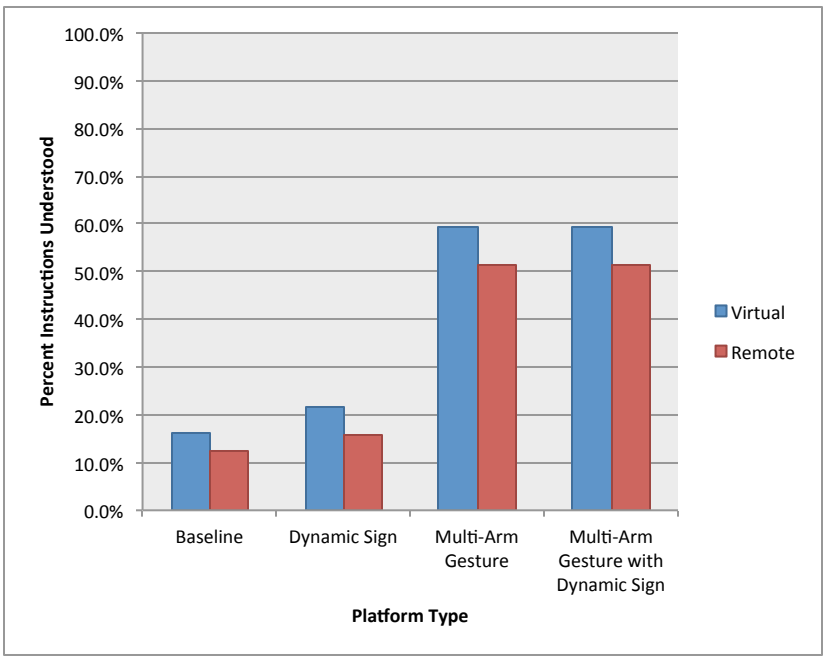


Figure 5.15: Results at the far distance level for the Remote Presence Experiment.

of 16) of participants interpreted the spinning motion that we intended to mean stay in place as an indication that they should turn around. Comments indicated that participants were unsure about their answers even though they could clearly see the robot. The far case presented additional difficulties for participants. They were generally unable to discern any identifiable motion from the robot. For each instruction at the far distance, four participants thought that they should follow the robot to the instruction point.

Dynamic Sign Again, a small difference was found between the remote and virtual platforms for the dynamic sign. Overall, 56.3% of instructions were understood on the remote platform compared to 60.9% on the virtual platform ($\chi^2(1, N = 64, R = 256) = 0.543, p = 0.446$). In the near case, where participants could definitely read the sign, 96.9% of responses indicated participants understood the instructions. We expected 100% of responses to indicate understanding, as in the virtual case, but one person answered that it was telling him to go right when it was actually indicating left (although the explanation given indicates the participant understood the intention) and another participant reported that one of the four videos would not play for technical reasons. Thus we can conclude that the dynamic sign is understandable at the near distance level. At the far distance level it was almost completely unintelligible by design and only 15.6% of participants answered correctly. Most participants indicated that they should follow it in all cases. They could

clearly see that the platform went down the hallway and turned to face them. They interpreted that action as the robot proceeding ahead and then waiting for them to catch up.

Multi-Arm Gesture The remote Multi-Arm Gesture platform performed about the same as the virtual platform. Overall, 65.7% of instructions were understood for the remote platform compared to 71.1% for the virtual platform ($\chi^2(1, N = 64, R = 256) = 0.581, p = 0.347$). Participants understood a majority of the instructions in the near videos. There was some confusion as to whether the forward and backward instructions actually meant to follow the robot, but a large majority of participants understood the instructions. Five participants were unable to understand the stay instruction. One thought it was indicating "no," and thus to turn around and go backwards, by crossing its arms. The others answered "unknown" and indicated they had no guess. One reported that he thought the robot was panicking. Results are worse at the far distance level with only 51.6% of participants reporting that they understood the robot's instructions. Many participants could not discern understandable gestures at this distance and thus answered that the robot must be asking them to follow it down the hallway. This is similar to the responses in the virtual study. In the virtual study, many participants thought the robot's arms were actually antennas or tentacles, but in this case only one participant seemed unclear that they were arms, referring to them as "flippers."

Multi-Arm Gesture with Dynamic Sign The remote Multi-Arm Gesture with Dynamic Sign performed closest to its virtual counterpart out of all of the robots tested. There was a 3.9% difference overall and a 0% difference in the near case, smaller than any other condition for any robot ($\chi^2(1, N = 64, R = 256) = 0.530, p = 0.477$). Comment responses were also very similar to the virtual case. When the dynamic sign was visible throughout the entire video (the near backwards and stay instructions) participants answered exactly as we expected. When it was obscured for a portion of the time (the near left and forward cases) a small number of participants were unable to understand the instructions (one in the left case, four in the forward case). At the far distance level the robot performed exactly the same as in the remote Multi-Arm Gesture case with very similar explanations from participants. Again, only one participant did not understand that the robot had arms, referring to them as "antennas."

5.3.3.2 Physical Presence Experiment

A total of 48 participants (denoted as N below, 30% female, mean age of 24.7 years old) responded to a total of 192 questions (denoted as R below) for this experiment over fourteen sessions. The

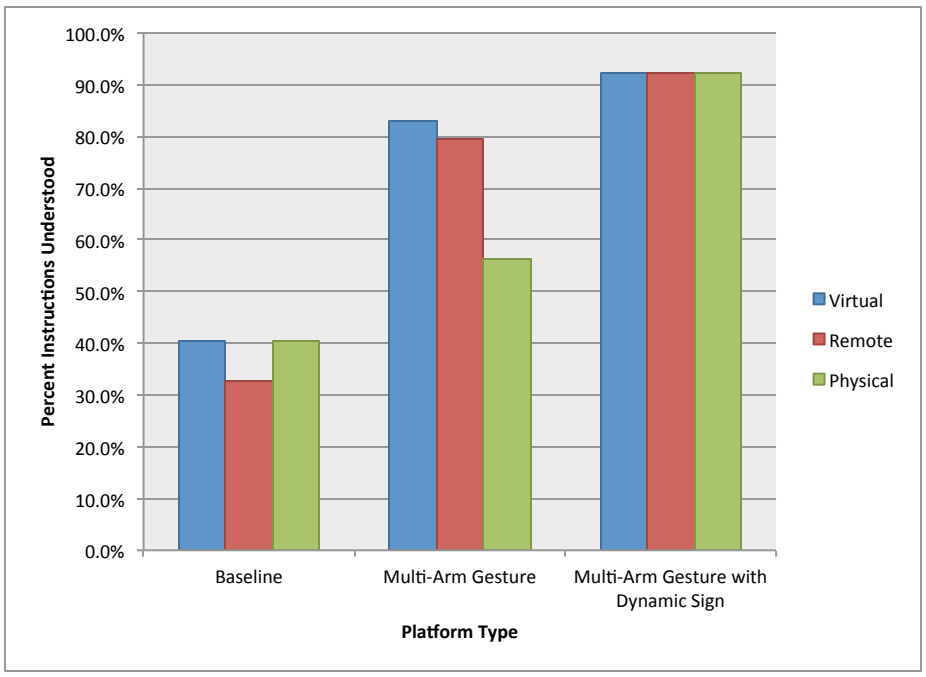


Figure 5.16: Results for the physical experiment compared with corresponding platforms in the virtual and remote experiments at the near distance level. Note that the Dynamic Sign platform was not tested in the physical experiment.

number of participants in each session ranged from one to seven. The results were broken up by platform type and compared to corresponding platform types at the near distance level in the virtual ($N = 48, R = 192$) and remote ($N = 48, R = 192$) experiments. Results can be seen in Figure 5.16. Participants viewed an actual robot performing live demonstrations, so there were occasional robot failures. Of the 56 total gestures performed, three were repeated. One due to the arms losing sync during the wave procedure (one arm was up and the other was down instead of moving together), one due to operator error (the wrong gesture was chosen) and one due to a participant arriving late. In each case, participants were instructed to ignore the failed demonstration and only answer survey questions about the correct one.

Baseline The Baseline platform showed no difference in the physical experiment when compared to the near condition of the virtual experiment (40.6% of participants understood the instructions) and a 7.8% greater understandability when compared with the remote experiment results at the near distance level ($\chi^2(2, N = 48, R = 192) = 0.191$, $p = 0.575$ across all three presence levels). No surprises were found in the comments, either. As in the previous experiments, participant comments generally indicated confusion rather than understanding for this platform.

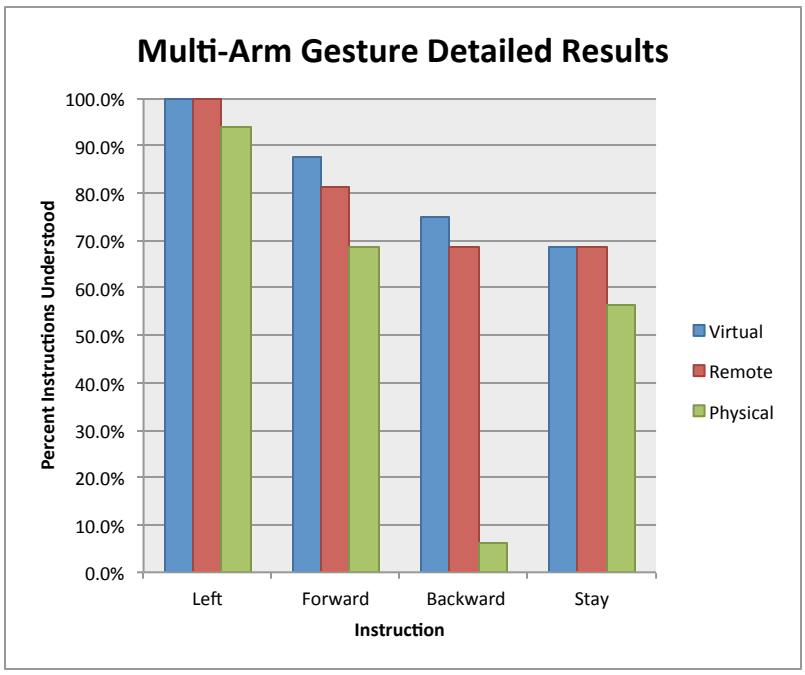


Figure 5.17: Detailed results of the Multi-Arm Gesture platform at the near distance level across all three presence levels. The only major anomaly is participants' inability to understand the "backward" instruction in the physical experiment.

Multi-Arm Gesture The Multi-Arm Gesture platform did show a significant difference based on presence level ($\chi^2(2, N = 48, R = 192) = 0.393$, $p = 0.001$). Over all instructions, the physical platform was 26.5% less understandable than in the virtual experiment and 23.4% less understandable than in the remote experiment. This was a curious result, so further analysis was warranted. Comparing responses to individual instructions across the three presence levels revealed that the biggest difference was in the understandability of the backward instruction (Figure 5.17). The other three instructions ranged from a 6.2% to a 18.7% difference between presence conditions (left: $\chi^2(2, N = 48, R = 48) = 0.274$, $p = 0.360$, forward: $\chi^2(2, N = 48, R = 48) = 0.254$, $p = 0.413$, stay: $\chi^2(2, N = 48, R = 48) = 0.142$, $p = 0.695$), but the backward instruction had a 62.5% difference between the remote and physical conditions and a 68.7% difference between virtual and physical ($\chi^2(2, N = 48, R = 48) = 0.393$, $p < 0.001$). We believe that this is entirely due to our experiment location. Recall that participants were lined up along a wall to observe the robot and that we had hoped the doors in the wall would provide a believable route in the backwards direction. Instead, participants indicated in the comments that the robot was pointing at them but that no route was available behind them and thus the robot must be telling them something else. Five believed that the robot wanted them to follow it, five decided on stay in place and four thought the robot actually wanted them to move forward, believing that this was a beckoning gesture.

Multi-Arm Gesture with Dynamic Sign The final platform performed exactly the same over all four instructions as it did in both the virtual and remote conditions at the near distance level. For each instruction, only one or two participants did not understand the direction correctly. One participant indicated that the robot gave two different directions during the backward instruction. This was recorded as an unknown in our results because the participant could not distinguish between the two directions. Both participants who did not understand the forward instruction indicated that they thought the robot wanted them to follow it, which is similar to the remote and virtual experiments.

5.3.4 Discussion

All of the results in the remote experiment and the physical Baseline and Multi-Arm Gesture with Dynamic Sign show a small difference between the virtual, remote, and physical experiments. These platforms ranged from a 3.9% difference to a 5.8% difference over all instructions and distance levels for the remote experiment and between 0% and 7.8% for the physical experiment. None of these results were significant at a $p < 0.05$ level. The only different platform condition, Multi-Arm Gesture in the physical experiment, only had a significant difference in a single instruction. As explained above, we believe that is because the location of the experiment did not have an obvious exit route in the direction the robot indicated, and thus this result is spurious. Qualitatively, participants gave similar explanations for their interpretations of the instructions in all three experiments.

As in our virtual study, participants attempted to understand the robot's instructions with any information that they had. All participants gave some comments to explain their response. One participant tried to help our design process by suggesting we use colored lights and loud sounds to aid people with cognitive disabilities or people taking narcotic medication in understanding the robot's instructions. Many participants in the remote experiment observed that there is a green light on the back of the Turtlebot base. This light shows brightly in the video but is dim in person, so we had not considered it as a potential feature of the robot. Participants who noticed the light interpreted the green light as a signal to mean go forward or follow the robot. This was only observed in cases where no intelligible guidance instructions could be seen by participants, such as the Dynamic Sign Platform at the far distance level or the Baseline platform at either distance level. Any participants who could see arm movements, for example, ignored the green light and focused on the intended gestures. The same effect was seen when the robot would tend slightly to the left or to the right at the end of its path: participants would interpret this minor deviation in course as an indication that

they should proceed in that direction. Thus we can infer that if participants cannot understand the instructions of a robot they will attempt to glean knowledge out of any feature visible, no matter how insignificant or unintentional that feature was to the robot designer. No similar effect was found in the physical experiment because participants only observed near robots.

A valid study should have as diverse a population of participants as the expected population of future users. An emergency guidance robot could be used in many situations with many different populations, so we used crowdsourcing to gather a diverse population. When the demographics of our remote study population were compared with our physical study population, we found that the physical study population was much younger (average of 9.4 years younger) and had a somewhat lower female to male ratio. Moreover, 46 of 48 participants in the physical study indicated that they were students. This is not surprising given that recruitment for the study was mainly confined to the Georgia Tech campus, but many other studies use a similar recruitment strategy without attempting to gather a wider audience. For our study, this did not matter as participants in the physical and remote studies gave almost identical answers, but other studies may not be so fortunate.

5.4 Conclusion

This chapter has explored several strategies that a robot can use to communicate directional information to a human. We found that a simple mobile platform, even with distinct signage, is not sufficient to convey guidance information. A single arm has good visibility and can convey directional information but has difficulty in conveying other guidance information, such as "wait here". A dynamic visual display can show familiar pictograms and text instructions but has difficulty being seen over great distances. Two arms, whether on a humanoid robot or on a standard mobile base are capable of being seen at a great distance as well as communicating all necessary forms of guidance instructions. Thus, the following chapters use a robot with a mobile base and two arms for all future emergency scenarios tested. These results were first obtained by having participants evaluate videos of virtual robots, then by having participants view videos of physical robots and finally by having participants observe the actual robots in a laboratory setting.

Our results in this chapter show that there is a small difference between virtual, remote, and physical robot presence in HRI experiments that relate to understanding instructions given by robots. Only one platform had a significant difference in responses between the presence levels tested. We do not generalize this result to mean that all HRI experiments can be performed in a virtual setting, but rather that this is one experiment in a subset of all experiments that can be performed in a virtual

setting. We feel confident that other experiments which rely on a participant's ability to understand a robot's actions would be valid using a virtual setting. Additionally, performing this experiment first in a virtual setting, then in a remote setting, and finally validating in a physical setting allowed us to generate a number of virtual robots, prune these to a few useful physical designs and then use crowdsourcing to again prune our platforms and experimental conditions before performing a costly and time-intensive physical presence experiment. We believe other design processes can benefit from a similar process when developing new robots for HRI tasks.

In this chapter, we have established that emergency guidance robots can be understood in controlled environments, but will evacuees trust the robots to lead them to safety? The next chapter discusses several experiments that tested human-robot trust in time-critical situations.

Chapter 6

Factors that Impact Human-Robot Trust in Emergencies

6.1 Introduction

In the previous chapters, we have shown that robots can assist evacuees in an emergency as well as explored the methods that a robot can use to provide guidance instructions in these situations. In order to develop robot behavior that can modify a human's trust level, we must first understand what factors cause an individual to trust or not trust a robot in an emergency situation.

Risk, as defined in [78] and applied to the emergency domain, can be interpreted as a combination of situational risk and agent risk (Figure 6.1). Situational risk is the amount of danger that the victim perceives in the environment around him. Triggered fire alarms and the presence of smoke

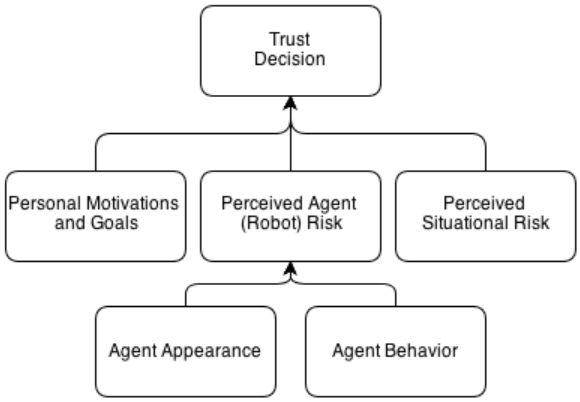


Figure 6.1: Factors that affect a human's trust in a robot

would increase the risk in a fire emergency. The sound of gunshots would increase the risk in an active shooter scenario. Very little risk might be perceived if there is no visual or audio indication of an emergency. Increased risk in the environment should generally increase the likelihood that the victim will follow the robot's directions in most situations.

Agent risk is considered in terms of both the agent's behavior and appearance. In this case, the agent is the robot trustee attempting to help the victim. In [58] we outlined the basic requirements for the appearance of an effective evacuation guidance robot. Following these guidelines will help to increase trust in the robot. In [61], we examined which actions the robot could perform to increase trust in situations where evacuees have personal reasons to disregard its directions. Many behaviors may increase the perceived risk of following the robot, such as the robot making the obvious error of colliding with an obstacle. There may be a perceived risk of the robot itself harming the victim. An increased perceived risk of following the guidance of the robot will generally lower the trust in the robot. In this chapter, we first describe our initial studies in human-robot trust, then present a study that determined how two risk factors affected the trust a human has in a robot for our third contribution:

Measured the effect of risk modality and robot effectiveness on human-robot trust.

6.2 Developing Methods to Evaluate Human-Robot Trust

Unfortunately, few research protocols exist for investigating human-robot trust. The methods that do exist have largely focused on very narrow aspects of the trust phenomenon and/or situations [63, 33]. Further, by definition, the presence of trust implies risk on the part of the person or the robot [78]. Placing study participants at risk is challenging from an ethical point of view and presents logistic problems. For example, an experiment may ask participants to move around a building while a fire is simulated using artificial smoke, visible flame, and fire alarms. One participant may view this experience as completely artificial and thus feel no risk, while another participant may panic and injure himself or herself in the same situation. Moreover, measuring trust is strongly influenced by factors outside of the experimenters' control. These factors make the investigation of human-robot trust extremely difficult.

Our approach for handling these challenges has been refined over numerous different experiments and in this section we present the methodological findings from experiments involving 510 participants. Here, we present the lessons that we have learned over the course of conducting these studies with the aim of informing future human-robot trust research. A brief listing of the major

Table 6.1: A list of the major experimental milestones discussed in this section and related to our study of human-robot trust.

Type of Experiment	Name	Number of Participants	Measurement	Motivation for Participants
Narrative	Pilot Experiment 1	20	Agreement with definition	N/A
Narrative	Pilot Experiment 2	32	Agreement with definition	N/A
Narrative	Full Experiment	128	Agreement with definition	N/A
Single Round Robot Guidance	Trust Matrix Experiment 1 (Small Mazes)	30	Decision to use robot and self-reported trust	Monetary Bonus
Single Round Robot Guidance	Trust Matrix Experiment 2 (Large Mazes)	57	Decision to use robot and self-reported trust	Monetary Bonus
Single Round Robot Guidance	Series of Pilot Experiments (varied questions and outcome matrices)	64	Decision to use robot and self-reported trust	Monetary Bonus
Single Round Robot Guidance	Equal Outcome Matrix Experiment	59	Decision to use robot and self-reported trust	Monetary Bonus
Single Round Robot Guidance	Trust and Equal Outcome Matrix Experiment	120	Decision to use robot and self-reported trust	Time constraint in emergency scenario

experimental milestones discussed in this section can be found in Table 6.1 below. We began with experiments that used written narratives to explore trust situations and then expanded into experiments that asked participants to make trust decisions in single round simulations with guidance robots. We then applied these lessons to develop the two round experiment discussed in Section 6.3. Throughout these experiments, we tested a variety of metrics, motivations, and behaviors.

6.2.1 Crowdsourced Narratives in Trust Research

Crowdsourcing has become a popular method to increase the number and diversity of participants in human-computer interaction and even human-robot interaction experiments [43]. We chose to crowdsource our experiment in order to broaden the pool of people from which our data was generated. The greater the variety in our participant pool, the greater the generality of our results. Crowdsourcing uses the combined resources of a large group of people connected over the internet,

to accomplish a goal or perform a task. Studies have examined the use of crowdsourcing as a means for garnering experimental subjects and found that the validity of experiments utilizing crowdsourcing is not otherwise compromised [31]. In order to guarantee quality work, only Mechanical Turk workers with overall acceptance rates 95% and above were used. To ensure diversity no participant was allowed to enroll in the study more than once. Workers that attempted to enroll in the study more that once were warned that their data would be rejected and pay refused. We also rejected responses that included incomplete answers and comments. This research was approved by the Georgia Institute of Technology Institutional Review Board.

6.2.1.1 Trust Definition Validation

Our initial research goal was to evaluate our definition for trust (Section 2.3) and the four conditions derived from this definition. To accomplish this goal, we needed a clear and understandable way to present different matrices to participants. We decided to use textual narratives (i.e. stories) as a way to present the matrices in a manner that most people could understand. We felt that narratives allowed a great deal of flexibility for creating situations that closely matched the original matrix. Moreover, the use of narratives only required basic reading skills in order to participate in the study. Finally, because outcome matrices are often described as short stories (e.g. prisoner's dilemma, stag hunt game) the use of narratives was a natural fit.

In order to empirically evaluate Wagner's conditions for trust (Section 2.3 and [78]), we needed to create narratives that matched outcome matrices that met and did not meet the conditions. We were able to further divide the matrices that violated the definition of trust into subcategories based on the way the definition was violated. For instance, a matrix that contains equal outcome values did not put the trustor at risk and hence violates our definition for situational trust. Table 6.2 depicts the different matrix types. The first matrix in Table 6.2 represents a situation that requires trust and meets our conditions for trust. The other four matrices violate at least one condition of trust. The **Equal Outcomes** matrix violated all conditions by providing a situation where the trustor risked nothing in the interaction. The **Trustor-Dependent Trustee-Independent** matrix presented a situation where only the trustor's actions affected the outcome, thus the trustor was not placing any risk in the hands of the trustee. This violates the second and fourth conditions. Likewise, the **Trustor-Independent, Trustee-Dependent** matrix represents a situation where the trustor has no control whatsoever. If the trustor is not able to make a decision then the situation does not meet our definition of trust. This matrix violates conditions three and four. Finally, the **Inverted Trust** matrix presents a scenario where the trustor receives the worst reward when the

Table 6.2: The categories and descriptions of trust and no trust situations tested along with an example outcome matrix for each.

Category	Example	Description
Trust Matrix	Trustor: a_1^i, a_2^i / Trustee: a_1^{-i}, a_2^{-i} — \$2000, \$400 / \$0, \$400	Fulfills trust according to the definition and its conditions.
Equal Outcomes	Trustor: a_1^i, a_2^i / Trustee: a_1^{-i}, a_2^{-i} — \$2000, \$2000 / \$2000, \$2000	Violates all conditions of trust by removing all risk to the trustor.
Trustor-Dependent, Trustee-Independent	Trustor: a_1^i, a_2^i / Trustee: a_1^{-i}, a_2^{-i} — \$2000, \$0 / \$2000, \$0	Only allows the trustor to affect the situation. The trustor does not risk any outcomes on the actions of the trustee.
Trustor-Independent, Trustee-Dependent	Trustor: a_1^i, a_2^i / Trustee: a_1^{-i}, a_2^{-i} — \$2000, \$2000 / \$0, \$0	Only allows the trustee to affect the outcomes of the trustor. The trustor has no choice in the scenario.
Inverted Trust Matrix	Trustor: a_1^i, a_2^i / Trustee: a_1^{-i}, a_2^{-i} — \$0, \$400 / \$2000, \$400	Presents a situation where the trustor wishes for the trustee to break trust in order to receive the best outcome.

trustee intends to fulfill trust and the best reward when the trustee intends to break trust. Thus the trustor would wish that the trustee would act in a manner that breaks trust, rather than maintains it. This matrix violates the fourth condition.

Each participant was asked to read and evaluate twelve scenarios. Participants were paid $1.67 for the completion of their survey.

6.2.1.2 Iterative Development of Narrative Phrasing

The narratives that we created were based on several different scenarios that we felt offered some flexibility in terms of storytelling. One was an investment scenario meant to verbalize the investor-trustee game. A second scenario described a navigation task based on our interest in emergency evacuation. The final scenario was a hiring decision. The narratives were written to be as simple as possible while still allowing the flexibility to test each of our outcome matrices. The names Alice and Bob were consistently used to represent the characters in the scenario. The narratives began with a sentence or two introducing the scenario. Next, each of the 4 potential actions and their outcomes are described. The narrative ends with a statement describing the decision and resulting action that was taken by Alice or Bob and a question asking the subject whether or not they believed that the chosen action indicated trust. In order to rule out potential confounding factors, half of the narratives displayed a positively stated action and the other half displayed a negative action ("Bob chooses to hire Alice" versus "Bob chooses NOT to hire Alice"). In half of the narratives, Alice was the trustor and Bob the trustee and in the other half that was reversed. The ordering of the narratives, and the outcome amounts were all randomized. Participants were asked to explain each individual answer.

Best practices were used when developing the narrative surveys including the creation of several pilot studies, examination of within-subject reliability, use of randomization to eliminate biases, and measurement and evaluation of potential confounding variables. Figure 6.2 depicts the evolution of these narratives.

Not surprisingly, early pilot studies indicated that the wording of the narratives could influence participant decisions. This can be seen in Figure 6.3, where 86% of responses agreed with our definition when presented with a Trust Matrix, but only 49% agreed when an Equal Outcomes matrix was presented. For example, initially subjects were asked if the selection of an action indicated that one individual did not trust the other individual. Examining participants' explanations for their answers indicated that they generally understood the narrative and the actions taken by the trustor in the narratives, but some did not notice that we were asking about one individual NOT trusting

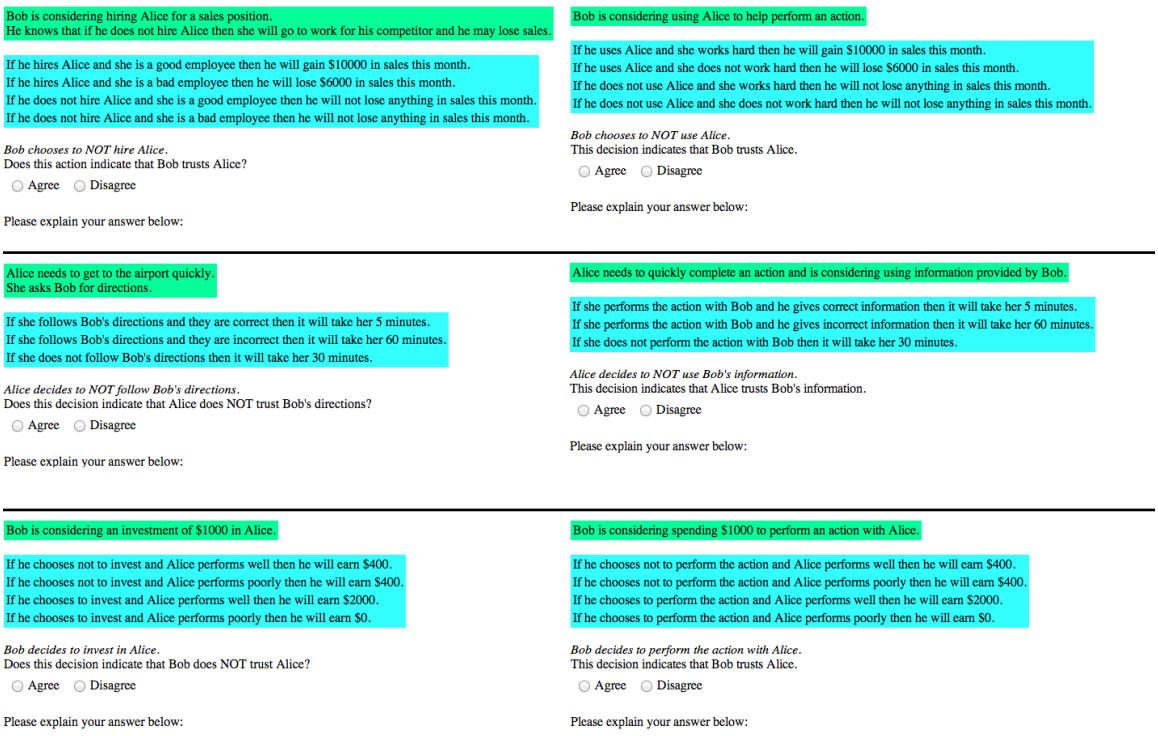


Figure 6.2: Initial iteration of the narratives (left) compared with their final version (right)

another. For some participants the negative phrasing led to confusion. We found that questions such as, "Does this decision indicate that Bob does NOT trust Alice?" could be interpreted in several ways. One interpretation is that trust is not involved or present during the situation. Another is that Bob distrusts Alice. Participants offered explanations such as "There was nothing for Alice to gain. So there was no need for her to trust. No distrust is indicated" and "It indicates neither trust nor distrust." After careful consideration, we eliminated the negatively stated trust questions believing that our working definition for trust and associated conditions could be adequately investigated with positive statements. This pilot study demonstrated that most individuals do not have clear delineations between notions such as "not trust", "distrust", "mistrust", and "trust is not required". Although our research is only interested in how people define "trust" rather than the various terms that indicate no trust, this may be a fruitful area of future research.

In an additional pilot study, some participants seemed to exhibit anchoring bias with respect to key words, such as "invest," "follow," and "hire" [75]. Anchoring bias describes the human tendency to focus heavily on early and/or specific pieces of information and disregard later information. This can be seen in Figure 6.3 where 93% of participants agree with our definition when a Trust Matrix is presented, but only 79% agree when an Equal Outcomes matrix is presented. Explanations by participants, such as, "In this case, even though the outcomes are the same regardless of Alice's

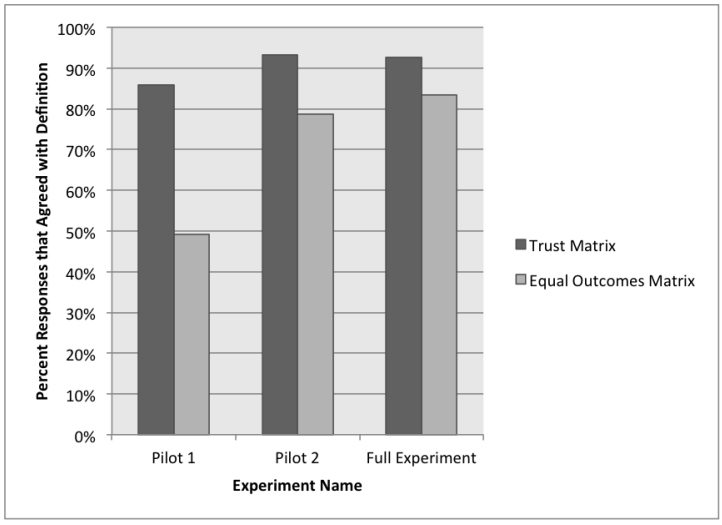


Figure 6.3: Results from the pilot and full experiments using textual narratives to describe potential trust scenarios.

decision, I would say that her choice to hire Bob is a sign of trust," "There is no situation where she 'loses' any money from either investing/not investing, she must believe that he can do good with the money," and "Since Bob decides to follow Alice's directions, this indicates that he trusts Alice. Though he will arrive at the destination regardless if he trusts her or not. If he knows this, it potentially makes it easier to trust Alice," clearly indicate anchoring bias. Because of this bias, we chose to replace specific actions that people were focusing on with less specific terms. For example, the statement "Bob is considering an investment of $1000 in Alice." became "Bob is considering spending $1000 to perform an action with Alice." The final iteration of the narratives used in this experiment removed all keywords, such as "invest" or "hire," and replaced them with less specific phrases, such as "perform the action." This allowed us to reduce anchor bias.

6.2.1.3 Results

In the full study, 128 participants' provided 1920 responses to the questions asked by the narrative. See Figure 6.3 for a comparison between this study's results and the corresponding results from the pilot studies. The scenarios showed minor, insignificant differences that appear attributable to random error. No significant difference regarding gender or magnitude of the outcome matrix values was found. The full results from this study are reported in [77], but some of our discoveries aided our later work. Overall, we found a strong correlation ($\phi(128) =_{0.592,}\ p < 0.01$) between the predictions of our conditions and the evaluations made by participants. Participants strongly agreed that the Trust Matrix narratives presented were indeed situations that required trust (93% agreement over

640 responses) but had some disagreements about situations that did not require trust according to our definition (66% agreement over all 896 responses for designated no trust scenarios).

At times, participants seemed to be confused by unusual situations described in the narratives, such as when Bob would perform an action with Alice even though doing so would cost him money or time. In these cases, some participants invented reasons that the trustor would choose or not choose to perform the action in order to make sense of a situation. Based on their comments, this appears to have occurred when they were confronted with a narrative that did not make sense. For example, when confronted with a situation where Bob decides to lose $2000 by participating in an action with Alice, one participant explains, "Bob trusts Alice because his decision has nothing to do with the money just his friendship with Alice." There is no mention in any of the narratives about a friendship or past relationship between the agents, yet the participant believes that there must be some reason Bob has chosen to lose this money and thus provides additional details so that the situation makes sense. With respect to the data, these peculiar narratives appear to have influenced participant trust evaluations more when the matrix did not meet our conditions for trust and may hint to a limitation of the use of narratives.

6.2.1.4 Discussion

Overall, the use of crowdsourced narratives to examine trust offered several advantages and disadvantages. Advantages include the ability to reach a large and diverse population of subjects, flexibility in terms of describing trust scenarios, and an ability to develop narratives that closely matched the matrices from which they were derived. For example, it is difficult to examine the trust involved in a hiring decision without using some type of narrative. Because we believe that our framework can be used to represent most situations involving trust, it was important to capture results from several different scenarios. This approach is not without its limitations: it was difficult to manage or eliminate all psychological biases, the narrative approach was disconnected from our larger goal of exploring human-robot trust, and translating these matrices into narratives resulted in some unusual descriptions of situations.

While this study did not involve robots, it showed that we can use crowdsourcing to examine participant views on trust situations. Most importantly, the results validated Wagner's definition of trust [78], which allows us to use the definition in our human-robot trust research later in this chapter. The study gave us insights into the exact wording we should use to ask about trust and cautioned us that participants will invent additions to whatever story we tell them, if the story does not seem to make sense.

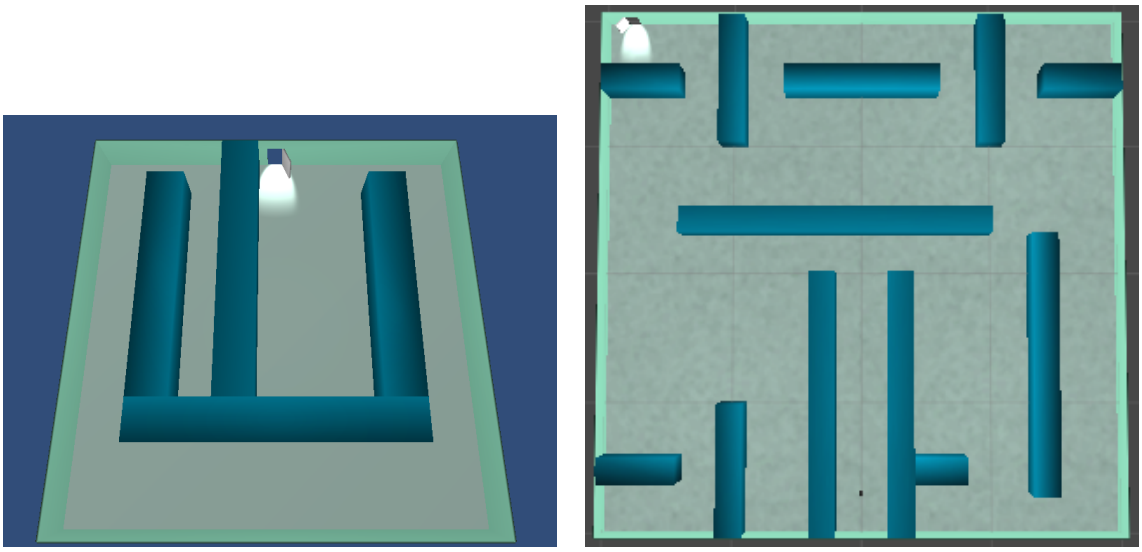
6.2.2 Single Round Evacuation Robot Experiments

As mentioned above, a key disadvantage of the narrative approach to investigating trust is its disconnection to robotics. In this section we describe experiments designed to test trust using an environment designed for human-robot interaction. Because a diverse set of participants was desired, crowdsourcing was once again utilized as a means for recruiting and paying study subjects. We developed a simulator that allowed participants to interact with a virtual robot using a web browser. For this experiment, participants were asked to choose whether or not they would like to use a robot for guidance when evacuating from a building. The building environment was modeled after a maze with corridors, dead ends, and no visual landmarks. Each simulation used the Unity 3D game engine to simulate the virtual maze and the virtual robot. Three-dimensional models for the game engine were created in Blender and Unity 3D. Participants were paid between $1 and $2, depending on the exact study. This setup is also used in the next section where we determined the effect of robot behavior and situational risk on a participant's decision to use a robot in a second round through the maze. Below is a brief overview of the single round experiments.

6.2.2.1 General Experimental Setup

Each experiment began by thanking the person for participating in the experiment. Next the subject was provided information about the evacuation task. In some experiments this included presenting the environment and robot to the subjects in videos, images, or text and providing information that allowed the participant to evaluate the risk associated with choosing to follow the robot. Participants were shown examples (again in the form of videos, pictures, and/or text) of good and bad robot performance (e.g. robots that are fast and efficient and robots that are not) and participants were given an idea of the complexity of the maze. Also, as part of this introduction, participants were given the chance to experiment with the controls in a practice environment. The practice environment was a simple room with three obstacles and no exit.

After this introduction, participants were given the choice to use the robot or not. With the exception of two pilot studies, participants were told that their choice to use the robot would not affect their compensation for this experiment. Participants were then placed at the start of the virtual maze. If they chose to use the robot it would start out directly in front of their field of view and immediately begin moving towards its first waypoint. The robot would move to a new waypoint whenever the participant approached. If the participant elected to not use the robot then no robot would be present and the participant would have to find the exit on his or her own.



(a) Example of an initial maze environment

(b) Example of a revised maze environment

Figure 6.4: Comparison between an initial maze environment and a revised maze environment.

After the maze-solving round was complete, participants answered a short survey about the round and about themselves. The exact questions asked in the survey varied considerably over the course of developing the pilot studies and the final experiment. This iterative process is described in Section 6.2.2.3.

6.2.2.2 Iterative Development of Scenario

Our first lesson in developing an evacuation simulation to test trust involved the size of the maze environment. In our first experiment, the maze environment was simple (see Figure 6.4a), so participants believed that they could easily solve the maze with or without the robot. As a consequence, their decision to use the robot was arbitrary, rather than based on trust, and they reported that on surveys. In response, we developed a maze environment sufficiently complicated that participants would have to rely on the robot in some way, if they chose to use it (Figure 6.4b).

Next, we attempted to create a representation of each of the outcome matrices in the narrative experiment (Table 6.2) in our simulator. This proved difficult. We developed numerous introductions to the experiment that informed participants that their monetary bonuses would be affected by choosing or not choosing to use the robot (i.e. Trustor-Dependent, Trustee-Independent). Additionally, we developed introductions that informed them that the robot would decide their bonus, regardless of their actions (i.e. Trustor-Independent, Trustee-Dependent). Finally, we told them that their bonuses were already decided and nothing in the simulation would have an affect (i.e.

Equal Outcomes). Very few participants read and understood these introductions. The few who did believed that they were being tricked by the experimenters and that their actions or the robot's actions would have an affect on their bonus. We focused on the Equal Outcomes matrix and, inspired by our first, overly-simple maze environments, found that we could inform participants that there was no risk in the scenario by removing the maze from the simulation environment. Participants could clearly see the exit at the start. We also informed them of this beforehand, so there was no illusion of risk at any time in the experiment. The results from this experiment are presented below.

6.2.2.3 Asking about trust

Our initial experiments found that participants would occasionally act as if they trusted the robot while reporting that they did not. This led to us to closely examine our method of asking about trust. Initially participants were asked: "When you made your decision to follow or not follow the robot, did you trust the robot as a guide in this scenario?" This produced good results when a Trust Matrix was used to design the experiment, but mixed results in other cases. Pilot studies were performed immediately afterwards, focusing on the Equal Outcomes matrix (see Table 6.2) and the trust question. We analyzed what it means for the participant to answer these questions. It was not initially clear if participants were stating whether they trusted the robot, the robot's ability to lead them to an exit, or something else. Additional pilot studies were performed with different wordings of the trust question. For example, we asked, "Did you trust the robot?", "Did your decision to follow or not follow the robot indicate that you trusted the robot?" and also varied responses available to the participants to include the option "Trust was not involved in the decision", in addition to "Yes," and "No." Overall, we found very little difference in the data resulting from these changes in wording.

In later single-round experiments, the issue of trust question wording was revisited. This time, participants who chose to use the robot were asked to agree or disagree with the statement "My decision to use the robot shows that I trusted the robot." Participants who chose to not use the robot were asked, "My decision to not use the robot shows that I trusted the robot." Each group was also asked if they trusted the robot itself. We again found very little difference in responses. Ultimately, we concluded that the wording of the question itself did not matter when compared with changes we made to the scenario.

6.2.2.4 Results and Discussion

After several experiments with varying motivations and simulation environments, we found a strong correlation between participant responses as to whether a situation required trust and our definition

of trust ($\phi(120) = +0.406, p < 0.001$). When conditions for trust were met, 74.0% of participants indicated that they trusted the robot, compared with only 32.9% when conditions were not met.

In the single-round experiments, we again validated the definition of trust but we also validated the use of robot guidance in mazes to test human-robot trust. We determined that the mazes used in these experiments were sufficiently complicated to present a trust situation and that people who chose to use the robot felt that their decision meant that they trusted the robot. In the next section, we again use robot guidance through a maze to test the effect of prior robot performance and the effect of situational risk on a participant's decision to trust a robot.

6.3 Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations

To develop trustworthy robots, we must first examine the conditions that affect a human's decision to trust a robot. One condition is prior task performance. In this section, we ask: how does the initial performance of the robot during a high-risk, time-critical situation affect a participant's decision to trust the robot later? The understanding gained by exploring this question will allow researchers to create robots that humans are more likely to trust, develop robots that understand how to better manage a person's trust, and may provide insight into the phenomenon of trust itself. To answer this question, we have developed an interactive navigation simulation that allows participants to use a robot as a guide to find the exit of a maze in a timed scenario. We measure the participant's decision to use the robot in an initial round, when the participant has little knowledge of the robot, and in a second round, after the participant has experience with the robot. We vary the behavior of the robot in the first round to determine the effect of successful and unsuccessful guidance on the participant's second choice. Two different methods were used to add time pressure to the scenario: a monetary bonus for a quick exit and a survival risk for not evacuating within a specified time.

6.3.1 Hypotheses

In order to explore how a robot's initial performance affects a person's trust, we must measure the change in trust after the robot acts as a successful guide and after the robot does not act as a successful guide. Our first hypothesis examines this question directly: (H1) *Self-reported trust will be significantly lower in the second round if the robot did not perform well in the first round.*

There are many ways for a robot to fail during a time-critical situation. For this guidance

scenario, one failure mode is for the robot to be an inefficient or slow guide. This occurs when the robot successfully leads the person to the exit, but requires a great deal of time to do so. Another type of failure is for the robot to not lead the person to the exit. One way to implement this type of failure is for the robot to stop moving somewhere within the maze. We hypothesize that: (H2) *Participants that are guided by a robot that fails will self-report less trust than participants that are guided by a slow, inefficient robot.*

As stated above, different measures for trust exist. One could use a measure of the person's behavior to infer the amount of trust. Alternatively, one could ask participants to self-report their trust. We hypothesize that: (H3) *There will be a high correlation between participants who decide to use the robot in a round and participants who self-report that they trusted the robot.*

Risk is a major component of trust [78]. Characteristics of the experimental scenario can influence a subject's perceived risk differently. For example, the risk associated by losing $10 gambling will likely impact the behavior of people near poverty more than wealthy people. From an empirical point of view, we would like to control the factors that influence the subject's perceived risk. Yet, monetary incentives are a common method for putting a person at risk in order to explore trust [42]. Our final hypothesis examines the use of emergency scenarios as a possible replacement for monetary incentives in trust research. This leads to our final hypothesis: (H4) *The decision to use the robot for guidance in the second round is significantly more sensitive to the robot's performance in the emergency scenario than in the bonus scenario.*

6.3.2 Methodology

To address these hypotheses, two different experiments were conducted. Both experiments required a person to navigate a simulated maze with or without the help of a robot. The person was required to navigate a different maze in two separate rounds in order to examine the impact that a robot's initial performance has on later decisions involving trust. They were given the option to use a guidance robot prior to navigating both mazes. Data reflecting their decision to use or not use a robot as well as surveys focused on the participant's reasoning were collected and used to confirm or refute the hypotheses presented above.

6.3.2.1 Participant Inclusion and Exclusion Criteria

Emergency guidance robots could potentially aid a large variety of people. In order to gather such a large variety of participants, crowdsourcing (via Amazon's Mechanical Turk service) was used

to collect data for both experiments. In order to ensure the best possible data, participants were required to have a 95% acceptance rate for their previous work and were only allowed to participate once.

The experimental surveys required subjects to comment on the reasoning behind their decisions. Much of our previous work has indicated that participants understood our questions and thought logically about the answers (see Chapter 5). A participant's data was excluded if comments were missing, nonsensical (e.g. if the comments were not understandable), or repeated throughout. Human participation in our experiments was approved by the Georgia Tech Institutional Review Board.

6.3.2.2 Experimental Protocol

The same general experimental setup was used for both experiments (Figure 6.5). Participants began each experiment by accepting a request on Mechanical Turk and clicking a link to a Unity 3D Web Player executable. Some participants had to download the Unity Web Player plugin to perform the experiment. Next, they viewed an introductory message that described the navigation task they were to perform. This page included photos of an exit and the guidance robot. The guidance robot varied in the two experiments. They were then offered the opportunity to practice navigating in a maze. They had a first-person view of the maze and used their keyboard arrow keys to move. After the practice session, they were presented with illustrative examples of human-robot performances in the maze. The nature of these examples varied with respect to the experiment. The examples impressed a particular outcome matrix onto the participant in order to give him or her some background knowledge on the expected behavior of the robot. The participant was then asked to decide whether or not they would like a robot to provide guidance during the first round of the experiment. After making their choice the person then navigated the maze and completed a short survey (Table 6.3). The survey collected qualitative and quantitative information about a participant's trust in the robot as well as comments on their decision to use the robot or not. They were then offered another opportunity to use the guidance robot in the second round. Next they navigated the maze in the second round and completed a short survey about their second round decision. Unknown to the participant, the robot's guidance performance in the second round always matched its performance in the first round. The experiment concluded with a final survey that collected demographic information about participant age, gender, country of residence, occupation, and education level. This survey also asked if participants have worked with a real robot in the past.

This research is motivated by the desire to better understand how people react to emergency

Table 6.3: Survey presented to participants after each round.

Number	Question
1	Did you choose to use the robot in the previous experiment?
2	Did you trust the robot?
3	Did you believe that the robot would find the exit quickly?
4	Were you motivated to find the exit as quickly as possible?
5	My decision to use the robot shows that I trusted the robot.

guidance robots so a simulation environment was created to resemble an office building. This environment included corridors and rooms designed to give it a maze-like appearance (Figure 6.6). Participants were placed in the environment with no prior experience and required to find a single exit.

6.3.2.3 Measuring Trust

The decision to use the robot was viewed as an indicator of trust. This decision served as a binary behavioral measure of trust: either the person trusts and uses the robot or the person does not trust and use the robot. Our conceptualization of trust focuses on the risk a person accepts when choosing to depend on the robot. Hence, we believed that the person's decision to use or not use the robot could serve as a measure of trust. In the first round, the participant must choose based on very little information, but in the second round the participant bases their decision on the robot's previous behavior. Thus, we felt that measuring the participant's decision to use or not use guidance from the robot at the beginning of the second round would provide a measure of their trust in the robot.

We also measured trust by asking participants to self-report whether or not they agree with the statement: "I trusted the robot when I made my decision to follow or not follow the robot." In addition to the options to agree or disagree, we offered the option of choosing "Trust was not involved in my decision." In pilot studies, we found that some participants felt that disagreeing with the trust statement meant that they actively distrusted the robot. We therefore, provided a third option that clearly indicates they neither trust nor distrust the robot. Our results are based on people that affirm their trust in the robot. The use of binary measures rather than Likert scores is common in trust research (e.g. [46]) and, we feel, more accurately reflects the types of high-risk decisions a person must make during an emergency.

6.3.2.4 Robot Behavior

The performance of the robot informs the human of the robot's ability to be trusted in future interactions. H1 examines how the robot's behavior affects the participants' self-reports of trust in

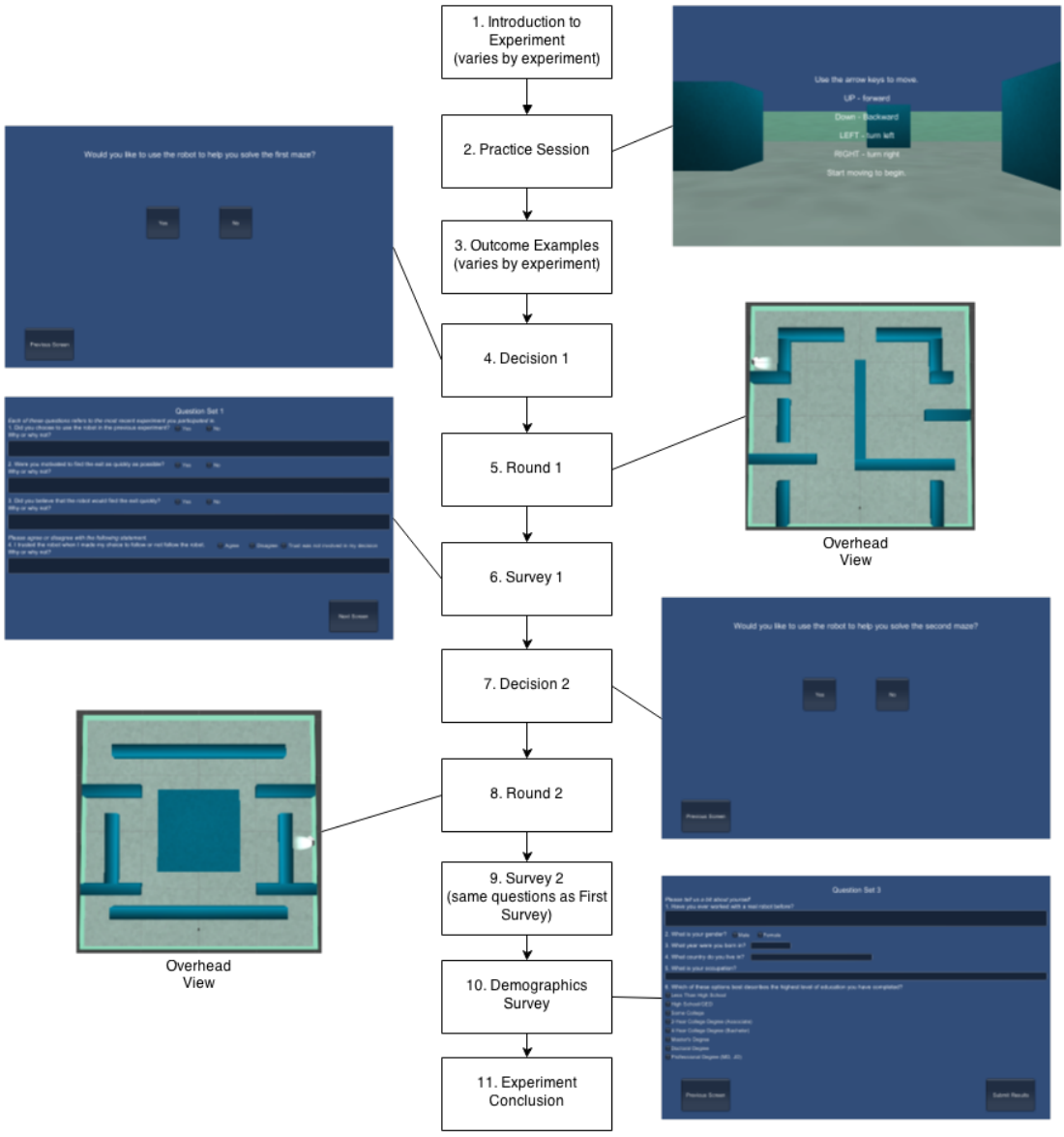


Figure 6.5: Experimental protocol with screenshots from experiment. The entire experiment was presented in a Unity 3D web game, including the survey questions.

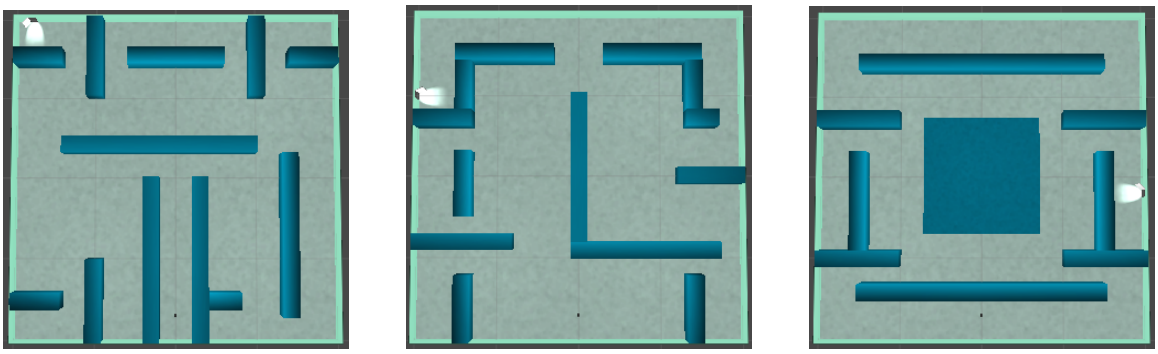


Figure 6.6: Overhead views of the three environments used in both experiments. Environments were designed to be similar to office layouts. Corridors and rooms were used to give maze-like qualities to make the simulation challenging.

the second round. H2 explores different types of robot guidance failures: one that inefficiently leads the person to the exit and one that fails entirely to lead the person to the exit. In pilot studies we evaluated several different types of robot guidance failures. All but two of these failure modes were eliminated because participants were unable to determine that the robot had failed and hence resulted in an extremely long experiment completion time (see Table 6.4 for a listing of the robot guidance failure types that were not included in these experiments). Overall, three robot behaviors were defined that were used in the experiments:

- Efficient navigation: the robot proceeds directly to the exit location (Figure 6.7). Robots that acted in this manner are capable of finding the exit within thirty seconds. This behavior was designed to preserve the monetary bonus in the bonus scenario and allow for a fast exit in the emergency scenario.

- Circuitous navigation: the robot explores many possible routes before eventually finding the exit (Figure 6.7). Robots that acted in this manner are capable of finding the exit in ninety seconds. This behavior was designed to garner small bonuses for participants in the bonus scenario and did not find the exit in time for participants in the emergency scenario.

- Incorrect navigation: the robot proceeds directly to a corner of the environment that is not the exit location and then stops. This is meant to emulate the behavior of a robot that has incorrect information about the exit location. Robots that acted in this manner stopped moving after approximately thirty seconds at a point at least thirty seconds from the exit. As in the circuitous navigation condition, this behavior did not allow participants to receive much, if any, bonus in the bonus scenario and did not guide participants to an exit in time for them to evacuate successful in the emergency scenario.

Table 6.4: Failed robot guidance behaviors that were used during a pilot study.

Name	Description	Reason for Exclusion
Small Loops	Robot circled an obstacle continuously	Several loops around the obstacle were required before participants realized the robot had failed. The total time for the experiment was too long.
Large Loops	Robot circled a large area of the environment continuously	Participants could not realize that the robot had failed until it completed at least one loop. This could take several minutes by itself and thus the total time for the experiment was too long.
Continuous Searching	Robot searched through entire environment except location of actual goal position. After completing a search it started again.	Participants followed the robot for considerable time before realizing the robot had failed. Some participants would follow the robot for more than 15 minutes.
Wall Collision	Robot nearly found goal but then continuously collided with wall and was unable to proceed.	Participants did not understand that the robot was colliding with the wall and thus did not understand that it failed.

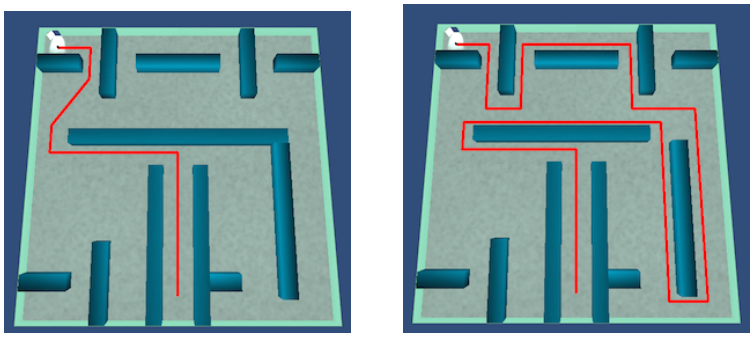


Figure 6.7: Examples of efficient robot guidance (left) and circuitous robot guidance (right). During efficient guidance the robot knows exactly where the exit is and effectively mitigates the participant's risk. During circuitous guidance the robot searches for the exit, eventually finding it.

The robot followed a predefined set of waypoints throughout the environment to perform the behaviors. Waypoints were set near corners or occlusion points so that the robot stayed in the participant's view. The robot waited at each waypoint for the participant to approach before it moved to the next waypoint. The robot was allowed to move considerably faster than the participant so that it would always be leading. The exact time to reach each end point depended on the particular environment and on the participant.

6.3.3 Experiment 1: Bonus Scenario

The first experiment examines the use of losing money as a way to put participants at risk. We used the risk of losing a potential allotted bonus as the source of risk motivating participants' trust

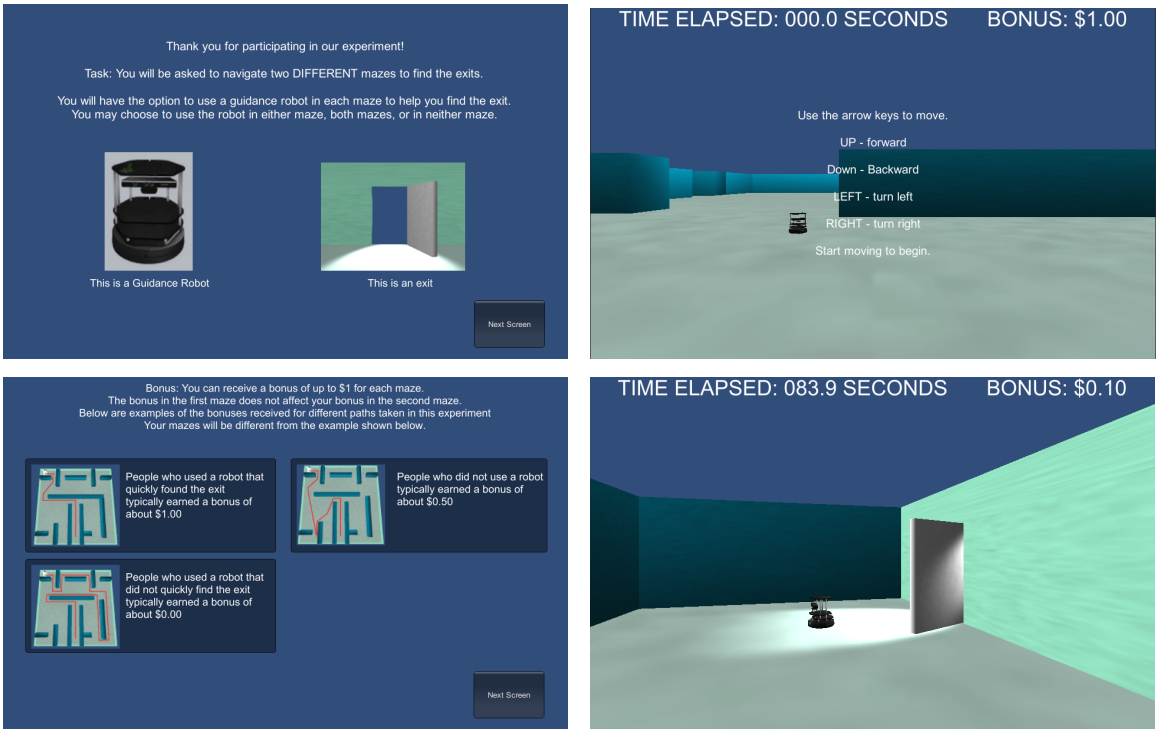


Figure 6.8: Screenshots from the bonus scenario experiment. The figure depicts the introduction screen (top left), example outcomes (bottom left), beginning of a round (top right), and successful navigation to an exit (bottom right).

decisions. This is an established procedure in the trust literature [42]. Subjects were offered a $1 bonus if they could find the exit of a maze within 30 seconds. After the first 30 seconds had elapsed the bonus began to decrease. Ninety seconds after the start of the experiment the bonus was $0. Participants were informed that their choice to use a guidance robot or not would not directly affect their bonus in any way.

The type of robot behavior (efficient, circuitous, incorrect) witnessed by the person served as the independent variable for this study. Measurements of trust served as the dependent variable. Both behavioral measures of trust and post-round self-reports were collected. The correlation between these two measures was used to evaluate H3. Hypotheses H1 and H2 were examined by comparing trust measures between subjects that interacted with different types of robot guidance behaviors.

6.3.3.1 Experimental Setup

Although the experiments followed the same general procedure described in Figure 6.5 above, some screens and text were unique for each experiment.

The first screen seen by the participants gave instructions. The simulated environments were specifically referred to as "mazes" to give the participant an idea of their complexity and goal. For

this experiment the robot displayed during the introduction and used in the rounds was a Willow Garage TurtleBot 2. The 3D model of the robot was created out of CAD files distributed by the manufacturer.

After the practice session, the participants were informed of the performance-based bonus and how to obtain it. Participants were given three example performances (Figure 6.8 bottom left) for the navigation task: (1) stated "People who used a robot that quickly found the exit typically earned a bonus of about $1.00" accompanied by a top-down view of a direct path to the exit in an example maze; (2) stated "People who used a robot that did not quickly find the exit typically earned a bonus of about $0.00" accompanied with a top-down view of a very indirect path to the exit in the same example maze; (3) stated "People who did not use a robot typically earned a bonus of about $0.50" accompanied with a top-down view of an indirect path to the exit in the example maze.

For this experiment, at the beginning of each round participants were informed that their bonus was currently set at $1.00 (Figure 6.8 top right). When the participant began moving, a timer in the top left of the screen displayed the time spent navigating to a tenth of a second precision. The bonus was prominently displayed in the top right corner. After thirty seconds of navigating the maze, the bonus began to decrease at a rate of $0.0167 per second (Figure 6.8 bottom right). The bonus was completely depleted after ninety seconds. The second round was setup the same as the first but with a different maze. All other aspects of this experiment proceeded as described in the Methodology section.

Because participants had no control over the amount of bonus they earned; they were all paid the full $2.00 bonus after their experiment was completed. This information was not made available to any participant before the experiment.

In this experiment we also asked one additional survey question in order to better understand choices for following the robot. Participants were asked to rate their motivations with respect to time, money, and enjoyment on a seven point Likert scale. They were then asked to rank these motivations in terms of importance from most to least important. The additional survey was only included to help design better experiments.

6.3.3.2 Results

A total of 106 participants (mean age=31.0, 60.4% male) completed the first experiment, 84.9% of which chose to follow the robot in the first round, with no prior knowledge of the robot's behavior. Figure 6.9 depicts the number of participants who used the robot in rounds 1 and 2 for the efficient and circuitous/incorrect robot behaviors and the self-reported trust in rounds 1 and 2 for the different

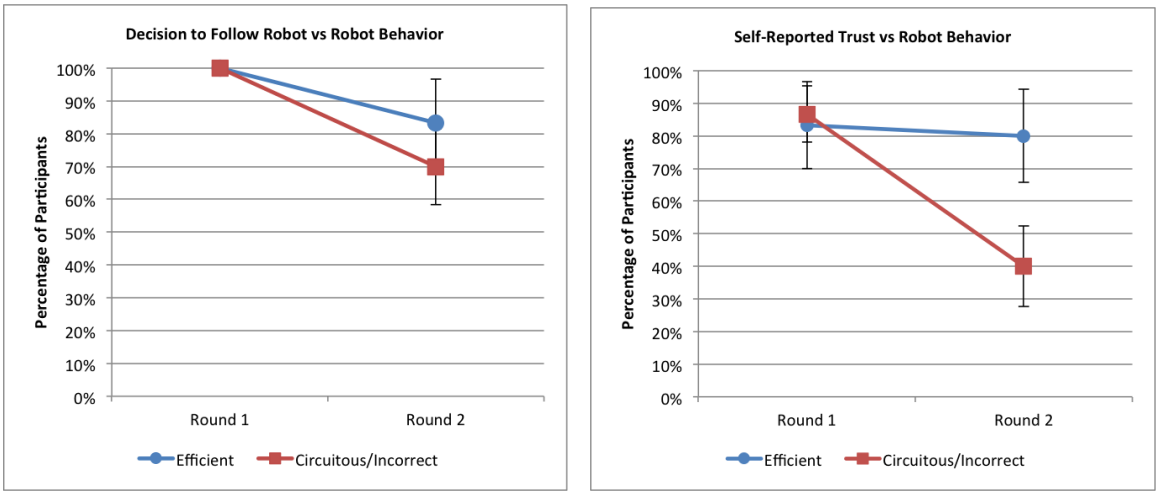


Figure 6.9: Change in decision to use robot (left) and self-reported trust (right) between the two rounds for the successful and unsuccessful robots. Note that a majority of participants continued to use the circuitous/incorrect robots even though half had lost their trust in the robot. Error bars represent 95% confidence intervals.

robot behaviors. Only participants who chose to follow the robot in round 1 are reported to ensure that all reported participants experienced the robot prior to their round 2 decision.

As can be seen in the figure, our first trust metric, the percentage of participants who self-report trust, decreases significantly (53%, $\chi^2(1, N = 60) = 68.76, p < 0.001$) when the participants experience a circuitous or incorrect robot in the first round. Only a 4% ($\chi^2(1, N = 30) = 0.11, p = 0.739$) decrease in trust was reported by participants that were guided by an efficient robot. There is a 40% difference in the level of self-reported trust in the second round between the efficient guidance behavior and the circuitous/incorrect behaviors ($\chi^2(1, N = 90) = 12.85, p < 0.001$).

The decision to use the robot (our other trust metric) did not see as large a difference. Efficient robots saw a 17% drop in decision to use the robot between the two rounds ($\chi^2(1, N = 30) = 5.45, p = 0.020$) while circuitous and incorrect robots dropped 30% between the two rounds ($\chi^2(1, N = 60) = 42.33, p < 0.001$). Still, there was only a 13% difference between the efficient robot usage and the other robots usage in the second round ($\chi^2(1, N = 90) = 1.87, p = 0.172$).

Figure 6.10 shows the results for the different failure modes. The type of robot failure had no impact on either the self-reported trust (0% difference) or the decision to follow (0% difference). In both the first and second round a strong positive correlation was found between following the robot and reporting trust in the robot, $\phi(106) = +0.628$ for round 1 and, $\phi(90) = +0.422$ for round 2.

We examined the survey comments to better understand each participant's rationale. Table 6.5 summarizes the most common comments from round 2. Note that, of the people that were guided

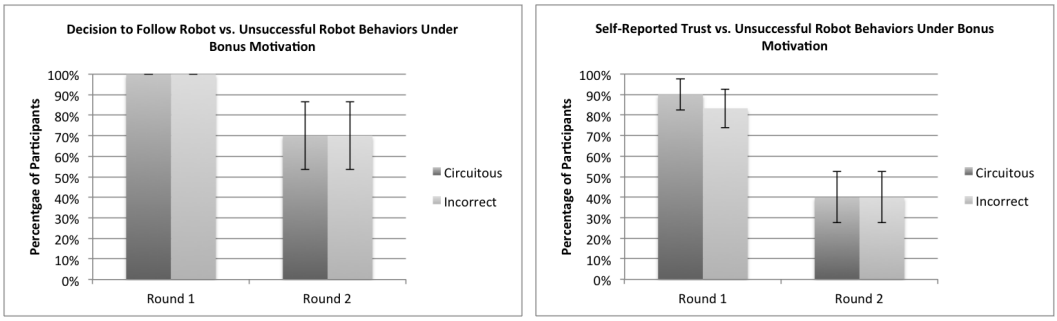


Figure 6.10: Change in decision to use robot (left) and self-reported trust (right) between the two rounds for the circuitous and incorrect robots. The same number of participants chose to use each and the same number reported trust in each in the second round. Error bars represent 95% confidence intervals.

by a circuitous or incorrect robot, many choose to follow the robot in the second round because they believed that the robot's help was better than no help at all (n=7) or they thought that the robot would perform better this time (n=5). These comments hint that participants were deciding to follow the robot in spite of the loss of bonus.

We performed an analysis on our motivational survey to better understand why people participated in our study. About half of the participants (55) reported that their most important motivation with respect to the experiment was money. The rest of the participants were evenly divided between time (25) and fun (24). These results indicate that many participants are not solely motivated by monetary bonuses in the experiment. Hence, some chose to follow the robot in the second round in spite of its failure and the fact that they self-reported not trusting it because they believed it would ultimately be faster or more fun to follow the robot.

6.3.3.3 Discussion

Overall, the results strongly support some of our hypotheses and do not support others. With respect to H1 the data indicates a 53% decrease in self-reported trust when the robot fails versus a 4% decrease when the robot does not fail. This result supports our hypothesis that self-reported trust significantly decreases after the robot provides slow or failed guidance. This result is important in that it shows that only a single failure can strongly and quickly influence a person's trust in the robot, which may have ramifications on the testing and evaluation of such systems. It is also noteworthy that the majority of people (84.9%) chose to follow the robot initially. This result appears to imply that people tend to trust initially when motivated by a monetary bonus.

Our second hypothesis focused on the manner in which the robot failed. We predicted that a robot that fails by traveling a short distance and stopping would have a significantly larger negative

Table 6.5: Summary of comments from Experiment 1.

Robot Behavior	Used Robot?	Self-reported trust	Comment Description
Efficient (n=30)	Yes (n=25)	Positive (n=22)	Robot performed well (n=21)
			Did not trust robot, trusted programmers (n=1)
		Neg./Neutral (n=3)	Impossible to trust machine (n=1)
			Trusted robot initially but explored on own instead of completing maze (n=1)
			More than two examples required to trust something (n=1)
	No (n=5)	Positive (n=2)	No complaint about robot, wanted to try experiment for themselves (n=2)
		Neg./Neutral (n=3)	No complaint about robot, wanted to try experiment for themselves (n=1)
			Thought robot would perform worse in second round (n=1)
Circuit. (n=30)	Yes (n=21)	Positive (n=11)	Robot performed better than human alone (n=7)
			Did not realize robot performed poorly (n=3)
			Thought robot would perform better in second round (n=1)
		Neg./Neutral (n=10)	Curiosity (n=6)
			Robot performed better than human alone (n=1)
	No (n=9)	Positive (n=1)	No complaint about robot, wanted to try experiment for themselves (n=1)
		Neg./Neutral (n=8)	Robot performed poorly (n=7)
			No complaint about robot, wanted to try experiment for themselves (n=1)
Incorrect (n=30)	Yes (n=21)	Positive (n=11)	Thought robot would perform better in second round (n=5)
			Did not realize robot performed poorly (n=3)
			Curiosity (n=3)
		Neg./Neutral (n=10)	Curiosity (n=6)
			Robot performed better than human alone (n=1)
	No (n=9)	Positive (n=1)	Unclear response (n=1)
		Neg./Neutral (n=8)	Robot performed poorly (n=8)

impact on both self-reported trust and the subsequent decision to follow than a robot that merely slowly led to the exit. The data shows that this is not the case. Our results indicate that either failure impacted trust equally both with respect to the self-report and subsequent following. The fact that there was 0% difference in both cases is presumably a statistical artifact.

Our third hypothesis examined the relationship between self-reported trust and the decision to follow the robot. We predicted that both the participant's trust and the likelihood of following would be strongly correlated. Indeed, we found a strong positive correlation between following the robot and self-reporting trust. Yet, the results show that numerous participants (26% of all participants) choose to follow the robot in the second round even though they reported not trusting it.

6.3.3.4 Experiment 1 Conclusion and Motivation for Experiment 2

The experiment supports our hypothesis that self-reported trust decreases after poor guidance. Further, we found a strong correlation between self-reported trust and the subject's decision to follow the robot. No statistically significant difference resulted from the different types of robot guidance failure. Still, the data indicates that participants were significantly more likely to follow the robot in spite of a failure and a lack of trust in the robot. An examination of the participant's comments hinted that the participants are motivated by factors *other than the bonus*, such as time and fun.

These survey comments and our study of the participants' motivations led us to develop a second experiment that sought to better align the participants' motivations with the task goals. This second experiment asked participants to act as if they were in an emergency. Instead of receiving a bonus, a quick exit from the building rewarded them with "survival." We hypothesized that being motivated by an emergency would result in a better alignment between their self-reported trust and their decision to follow the robot.

6.3.4 Experiment 2: Emergency Scenario

In a second experiment, participants were told that our goal was to discover how people leave a building during an emergency. Instead of receiving a bonus for a fast completion, they were told that they would only survive if they found the exit in time. During both rounds, a countdown timer appeared in the middle of their view to tell them the remaining time. As with the previous experiment, this study was conducted using the Unity simulation and Amazon's Mechanical Turk. Participants were compensated $2.00 for their participation in this experiment.

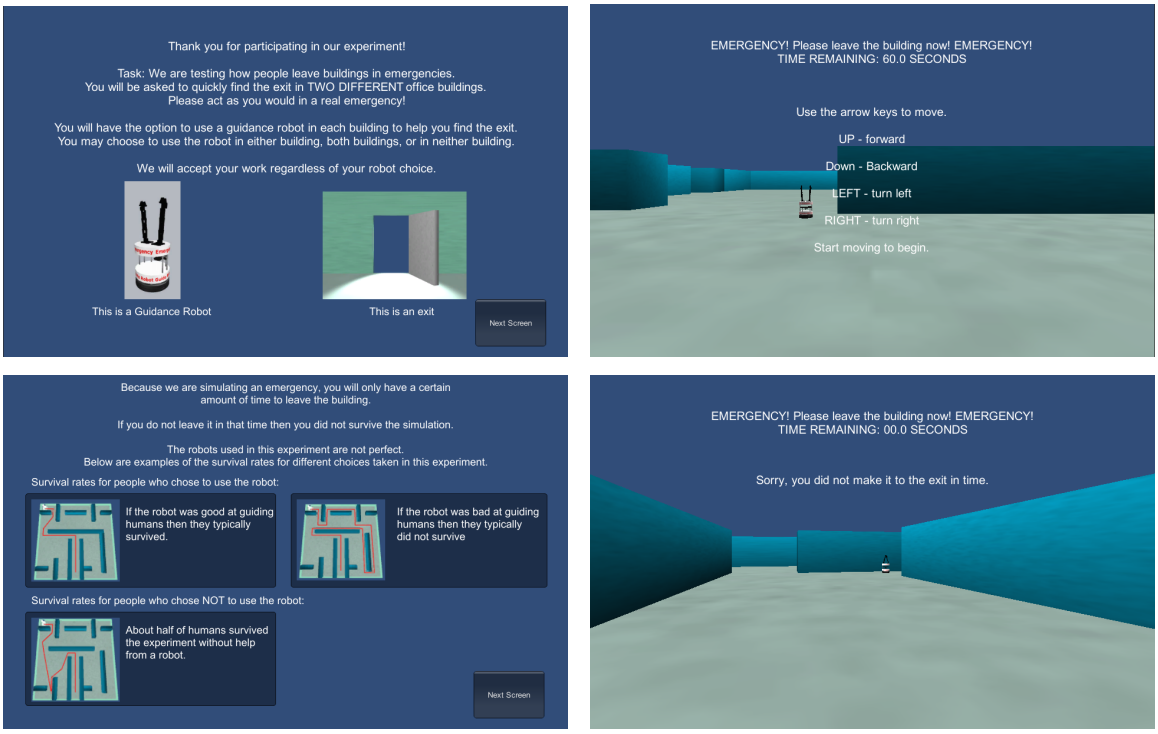


Figure 6.11: The introduction screen for the emergency scenario experiment is depicted in the top left. Note that the robot is different from in Bonus Motivation Experiment. Additionally, participants were told that this experiment was to determine how people evacuate buildings. The screen on the bottom left depicts example results. Participants were shown overhead views of the example environment with survival possibilities. The screen on the top right presents the beginning of the first round of the experiment. The timer counted down and was moved to the center of the screen for maximum visibility. Text indicated that an emergency had occurred. An example of an unsuccessful exit is presented in the bottom right. Text informed the participant there was no time remaining. The robot can be seen in the distance.

6.3.4.1 Experimental Setup

There were several differences between this experiment and the Bonus Motivation Experiment. First, the introduction screen stated "We are testing how people leave a building in emergencies" and asked them to "Please act as you would in a real emergency!" (Figure 6.11 top left). The word "building" was used instead of "maze" to further reinforce the emergency portion of the simulation.

The robot in this experiment was a TurtleBot 2 modified with two PhantomX Pincher AX-12 arms to allow it to gesture. The robot was also given signage to indicate that it is an emergency evacuation robot. The arms waved while it moved to attract attention. The robot's appearance and gestures were evaluated in the previous chapter (see Figure 5.3e) and it was found that participants understood it better than other forms of evacuation robots.

For this experiment each round ended after 60 seconds regardless of the participant's ability to find the exit. Once again, before selecting whether or not to use the robot, the participant was presented with a series of example experimental performances (Figure 6.11 bottom right): (1) stated "If the robot was good at guiding humans then they typically survived" accompanied by a top-down view of a direct path to the exit in an example maze; (2) stated "If the robot was bad at guiding humans then they typically did not survive" accompanied with a top-down view of a very indirect path to the exit in the same example maze; (3) stated "about half of humans survived the experiment without help from a robot" accompanied with a top-down view of an indirect path to the exit in the example maze.

During each round the words, "EMERGENCY! Please leave the building now! EMERGENCY!" appeared as well as the time remaining to exit (to a tenth of a second precision) in the top-center of the participants' view throughout the entire round (Figure 6.11 top right).

Other than these changes, both experiments were identical. Participants were again required to complete the same survey examining their trust in the robot and reasoning for choosing the robot for both rounds.

6.3.4.2 Results

A total of 129 participants (mean age=31.8, 60.5% male) completed the second experiment, 69.8% of which decided to use the robot in the first round. As shown in Figure 6.12 the decision to follow the robot decreases by 50% in the second round when the participant interacts with a circuitous/incorrect robot in the first round ($\chi^2(1, N = 60) < 0.01, p < 0.001$), compared to just 3% when an efficient robot is used first ($\chi^2(1, N = 30) = 1.02, p = 0.313$). There was a 47% difference in usage between

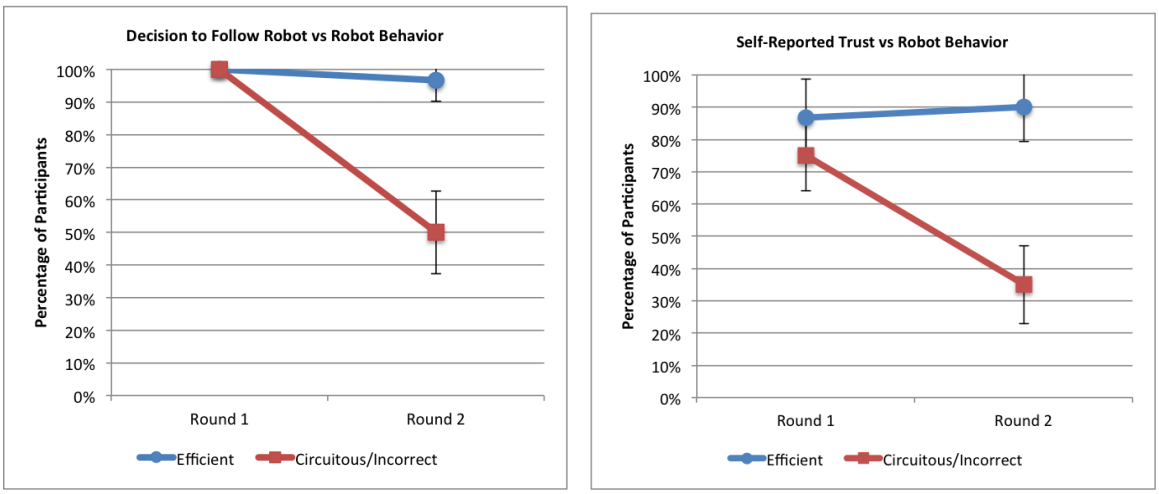


Figure 6.12: Change in decision to use robot (left) and self-reported trust (right) between the two rounds for efficient and circuitous/incorrect robots. Note that the decision to use the robot dropped with self-reported trust in this experiment, unlike in the Bonus Motivation Experiment. Error bars represent 95% confidence intervals.

the efficient guidance behavior and the failed behaviors ($\chi^2(1, N = 90) = 19.29, p < 0.001$).

Self-reported trust follows a similar trend with trust decreasing 53% when participants experienced a circuitous/incorrect robot ($\chi^2(1, N = 60) < 0.01, p < 0.001$) and self-reported trust increasing by 3% after interacting with an efficient robot in the first round ($\chi^2(1, N = 30) = 0.16, p = 0.688$). There was a 55% difference between efficient robot and failed robot trust levels in the second round ($\chi^2(1, N = 90) = 24.31, p < 0.001$).

Figure 6.13 shows the results for the different failure modes. The type of failure had minimal impact in the participant's decision to follow ($\chi^2(1, N = 60) = 0.27, p = 0.606$). There was also a minimal change in self-reported trust ($\chi^2(1, N = 60) = 1.15, p = 0.284$). A strong positive correlation was found between choosing to use the robot and reporting trust in the robot in both rounds: $\phi(129) = +0.661$ for round 1 and $\phi(90) = +0.745$ for round 2.

Again, motivations for participants' actions and reports can be found in the comments. A short description of a selection of these comments can be found in Table 6.6. Note that not all participants' comments are included in this table for brevity and some participants gave multiple reasons for their actions.

6.3.4.3 Discussion

The results from this experiment strongly support H1, H3 and H4 but do not support H2. A single failure of a robot caused 50% of participants to stop using the robot in the second round, compared to just a 3% drop with a successful robot. This supports our hypothesis that participants will

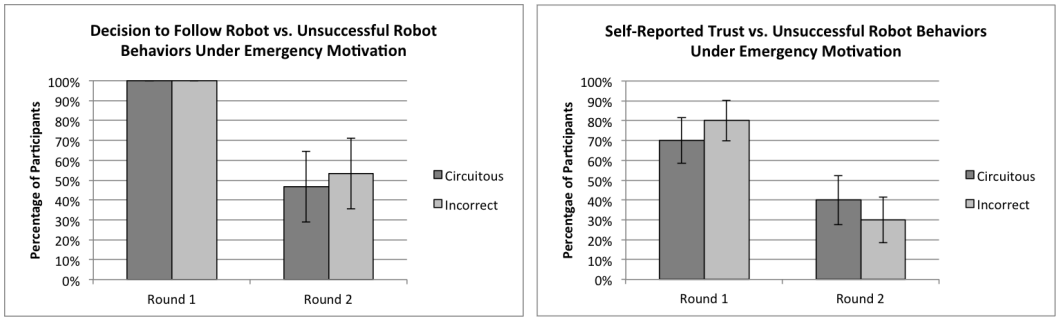


Figure 6.13: Change in decision to use robot (left) and self-reported trust (right) between the two rounds for the circuitous and incorrect robots. While the results are not identical in this round, as they were in the Bonus Motivation Experiment, they are still not statistically significant. Error bars represent 95% confidence intervals.

Table 6.6: Summary of comments from Experiment 2.

Robot Behavior	Used Robot?	Self-reported trust	Comment Description
Efficient (n=30)	Yes (n=29)	Positive (n=27)	Robot performed well (n=24)
		Neg./ Neutral (n=2)	Logical choice, not trust (n=1)
			Decided to proceed on own for fun after choosing to use robot (n=1)
	No (n=1)	Positive (n=0)	
		Neg./ Neutral (n=1)	Thought robot would perform worse in second round (n=1)
Circuit. (n=30)	Yes (n=15)	Positive (n=12)	Curiosity (n=5)
			Thought robot would perform better in second round (n=3)
			Robot moved quickly, and thus was trustworthy (n=2)
			Did not realize robot performed poorly (n=2)
		Neg./ Neutral (n=3)	Curiosity (n=3)
	No (n=15)	Positive (n=1)	Trusted robot to NOT find exit (n=1)
		Negative/ Neutral (n=14)	Robot performed poorly (n=13)
			No complaint about robot, wanted to try experiment for themselves (n=2)
Incorrect (n=30)	Yes (n=16)	Positive (n=9)	Robot performed better than human alone (n=6)
			Thought robot would perform better in second round (n=3)
		Neg./ Neutral (n=7)	Curiosity (n=5)
			Robot performed better than human alone (n=2)
	No (n=14)	Positive (n=0)	
		Neg./ Neutral (n=14)	Robot performed poorly (n=12)
			No complaint about robot, wanted to try experiment for themselves (n=2)

continue to trust a robot that performs well (H1).

Our data indicates that a smaller percentage of participants chose to use the robot in the first round when compared to the first round of the Bonus Scenario Experiment. While a majority still chose to use the robot, and thus our findings from previous work are still supported, we did not expect such a change. Many participants justified their choice by stating that they did not want to put their life in the hands of a machine. This indicates that people are more likely to initially trust a robot when there is a lower risk (e.g. a financial risk instead of a survival risk). This data serves as evidence that people take the emergency scenario, and the risk it entails, seriously.

With respect to the type of robot failure, both experiments showed no difference in either self-reported trust or the decision to use the robot if the person experienced a circuitous robot versus a robot that stopped moving before arriving at the exit. This is an interesting area for future work as it indicates that participants do not discriminate based on how the robot failed, only that it did fail.

The results from this experiment show an even greater correlation between self-reported trust and the decision to use the robot than was seen in the Bonus Scenario Experiment. This supports our third hypothesis: self-reported trust and the decision to use the robot are correlated. Only 12% of participants chose to follow a robot that they did not report trusting in the second round of this experiment.

We also found strong support for our fourth hypothesis: the decision to use the guidance information from the robot was more sensitive to the behavior of the robot in the emergency scenario than in the bonus scenario. This result suggests that an emergency scenario, in contrast to a bonus scenario, does influence participants to act in a manner that is aligned with their self-reported trust. Thus, we feel confident in our use of emergency scenarios to test human-robot trust. This is in contrast to much of the existing work that tests trust (such as [42]), which uses a monetary bonus to motivate their participants

The comments also indicate that participants took the emergency scenario seriously. Several comments note that individuals acted as if they felt real pressure to find the exit quickly (one participant wrote "It felt like a challenge, and I treated it as an emergency as instructed," another wrote, "Burning building, needed to get out"). Some likened it to getting the high score in a video game while others just wanted to "survive" the simulation. Participants who did not successfully survive the first round typically stated that they were upset with the outcome. Some were upset at their robot, some at themselves. Almost all participants who failed to survive in the first round vowed to live in the second. We believe these comments are evidence that using simulated emergency

scenarios fosters a sense of risk in the participant that is critical for human-robot trust experiments. This data serves as evidence that people take the emergency scenario, and the risk it entails, seriously.

6.4 Conclusion

In this chapter, we present studies involving 770 participants in order to validate our working definition and conditions for trust in human-human and human-robot interactions. In these studies, participants have examined outcome matrices, decided the extent to which the interactive situation described by these matrices demands trust, and informed us about the extent that situational risk and robot behavior affects their decision to trust. In the beginning, written narratives were used to sculpt these situations. Later, simulated evacuations through a maze were used to convey risk and force the participant to make a decision.

Overall, these experiments have taught us the following lessons related to the empirical evaluation of trust:

- It is difficult for a non-expert human to understand when a robot has failed at a task. For example, if the robot is built to be a navigation guidance robot, participants will expect it to be a good guide.
- Our results show that most people will initially trust an unknown robot.
- In these experiments, even a single failure strongly impacted a person's trust.
- We found that the manner in which the robot failed does not impact trust.
- In low risk situations (monetary bonus) people may act as if they trust the robot after a failure even if they self-report little trust; however, in higher risk scenarios (simulated emergency) participant's self-reports matched their decision to use the robot. Experiments which attempt to equate the person's risk to a bonus appear to underestimate other motivations such as time and fun.

Our pilot studies with various unsuccessful robot behaviors raise concerns about people blindly trusting robots to perform their stated task. While many people seem predisposed to not trust any new form of technology, others seem to instantly trust a new technology to perform its task, regardless of evidence to the contrary. Participants were willing to follow a robot in what we consider to be an obviously unsuccessful search for an exit for almost 12 minutes for $2 in compensation (the bonus had expired by this time). Even in our emergency scenario, 50% of participants who had

experienced an unsuccessful robot chose to use the robot again in a second round. The next chapter explores this topic in greater detail.

The experiments presented in this chapter provide understanding of the human-robot trust dynamic, but do not give us a complete picture of a person's interaction with an emergency guide robot. Each of these experiments was simulated in a virtual environment and the simulations were focused on finding the exit in a maze, rather than finding an exit in a building familiar to the participants. Participants were forced to either explore the environment on their own or rely on a robot, rather than a more realistic situation where they would have some knowledge of the building. In the next chapter, we examine human-robot trust in more realistic emergency scenarios, first with two experiments in a virtual environment to determine the effect of emergency exit signs and prior knowledge of the building on a person's decision to use a guide robot, then in a physical space where participants experienced elements of a real emergency.

Chapter 7

Emergency Guidance Robot Validation

7.1 Introduction

In the previous chapters, we determined that robot guides can be of assistance in emergency evacuations, we developed understandable methods of conveying this guidance to nearby evacuees, and we evaluated the factors that would affect a human's trust in the robot using virtual simulations of mazes. In the maze simulations, participants had no knowledge of the environment and thus had to rely on guidance provided by a robot or explore on their own. In real evacuations, people typically know something about the building, like the location of the front entrance, and they know how to find standard emergency exit signs to locate additional exits. In this chapter we ask the question: will people trust an emergency guidance robot during a real emergency?

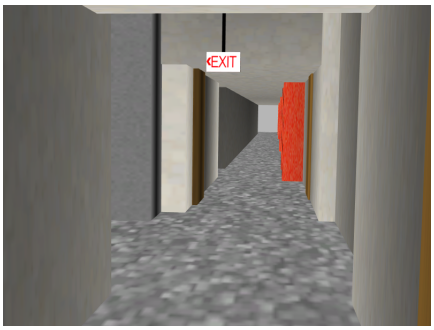
We began by giving participants the choice of following standard exit signs or one of the robots developed in Chapter 5 to find the exit in a virtual maze. We then created a virtual office environment and evaluated human-robot trust in a simulated emergency. In this simulation, participants had the choice of retracing their steps to the main entrance, following emergency exit signs to an alternative exit, or following the robot's guidance to an unmarked exit. Finally, we tested the same experiment in a building on campus, using artificial smoke and smoke detectors to simulate an emergency. In this experiment, participants had the choice of retracing their steps to the entrance of the building (also marked with an emergency exit sign) and following guidance from the robot to an unmarked exit. These experiments were performed in pursuit of our fourth contribution:

Measured a person's propensity to follow an emergency guidance robots in a realistic emergency scenario.

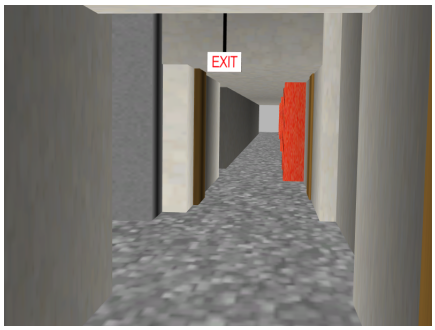
7.1.1 Verification of Exit Signs

We assume that most people are capable of understanding the intent of exit signs in the real world, but this may not be true of virtual exit signs in our 3D simulator. To test this, we performed a survey on Mechanical Turk similar to the surveys described in Chapter 5 in which we tested the clarity of our robots' guidance. Participants were asked to interpret four different images depicting different emergency exit signs (Figure 7.1). The first exit sign had an arrow pointing to the left from the participant's point of view and was intended to mean that an exit was located somewhere down the left hallway (Figure 7.1a). The second sign had no arrow (7.1b). This sign is often used to indicate that an exit is located somewhere down the hallway behind the sign. The third had an arrow pointing up and was intended to mean that an exit was located further along the hallway ahead of the participant (Figure 7.1c). Even though the second sign is standard in this situation, we provided this third sign to determine if an arrow is necessary in a simulated situation. The last sign displayed a different type of instruction and informed participants that this area is an emergency assembly point (Figure 7.1d). This was intended to be analogous to the "stay" instruction given by the robots in Chapter 5.

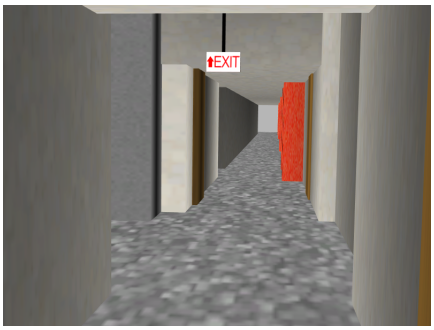
Sixteen participants took part in this survey. All participants saw each image. Results are presented in Table 7.1. Fifteen of sixteen participants (94%) were able to correctly identify the purpose of the first (Figure 7.1a) and the third (Figure 7.1c) signs (the signs with arrows). It is unclear why the last participant was unable to identify the directions depicted by these signs. Nine of the sixteen participants (56%) were unable to understand any guidance information from the exit sign without arrows (Figure 7.1b). Four participants (25%) correctly identified the sign as telling them to go forward, but the remainder were split between left (1), stay (1) and backward (1). We can conclude from this that most participants will require an arrow on an exit sign in a virtual environment in order to be comfortable with its instructions. Eight participants (50%) correctly identified the stay in place sign, even though the sign did not actually instruct them to take an action. Four (25%) stated they were unsure what the sign meant and the remainder were split between forward (2), left (1) and right (1). The two participants who indicated that the sign directed them to move forward did so because the sign was slightly in front of them, so they would have to move forward in order to be at the assembly point.



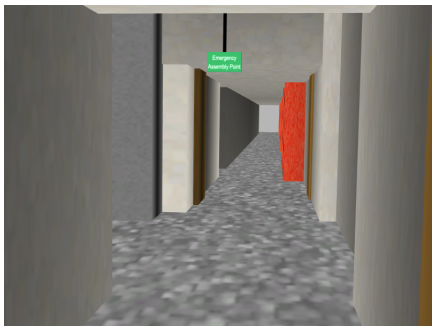
(a) Left exit sign



(b) Emergency exit sign without directional arrow



(c) Forward exit sign with arrow



(d) Emergency sign indicating participants should stay in place

Figure 7.1: Emergency exit signs shown to participants in our verification survey. Note that the simulated environment is the same as in Section 5.2.

Table 7.1: Participant understanding of static exit signs

Sign	Percent Understood
Left exit sign (Figure 7.1a)	94%
No directional arrow (Figure 7.1b)	25%
Forward exit sign with arrow (Figure 7.1c)	94%
Assembly point sign (Figure 7.1d)	50%

Based on these results, we used emergency exit signs with a directional arrow pointing to either the left or the right (as in Figure 7.1a) in virtual environments. This was one of the clearest signs in our survey and is a standard sign seen in many buildings.

7.2 Robot Guidance versus Existing Guidance Technology

In Chapter 5, we presented demonstrations of simulated (Section 5.2) and real robots (Section 5.3) to participants in order to evaluate the ability of various robotic platforms to provide understandable guidance information. That experiment provided useful feedback about the gestures themselves, but did not test human reaction to the robots in the context of an emergency. In Chapter 6, we tested human reactions in the context of an emergency, but we did not provide any additional aids to help a participant find an exit. In this experiment, however, we allowed people to experience a subset of the robots that we had previously tested in a 3D simulated environment during a simulated emergency. Participants were given the choice to follow guidance provided by a robot or guidance provided by emergency exit signs similar to those found in office buildings.

Evacuees exiting a building often encounter intersections which force them to make a decision about which direction to take. These decision points are usually accompanied by exit signs to help guide people to the closest exit. For this experiment, each decision point was also equipped with a robot to provide guidance. The guidance from the robot always contradicted the guidance from the exit sign. By measuring the person's choice at each decision point, we investigate the extent that participants trust robots more than signs, or vice versa.

7.2.1 Experimental Setup

Participants began the interactive portion of the experiment in a maze (Figure 7.2) facing a robot and a static emergency exit sign, one pointing left, the other right. Guidance information was presented at each decision point in the simulation. Five total decision points were used in the experiment. Two valid exits were available: one in the direction indicated by the robot and one by the sign. The participant had to follow the robot's or sign's guidance through at least three separate decision points to reach either exit, which limits the probability that a participant randomly chose to obey the robot or sign at all points.

The participant was encouraged to quickly find the exit. Prior to the experiment, instructions indicated that we were simulating an emergency. While navigating the environment, text on the screen stated "EMERGENCY! Please leave the building! EMERGENCY!" and a timer counted

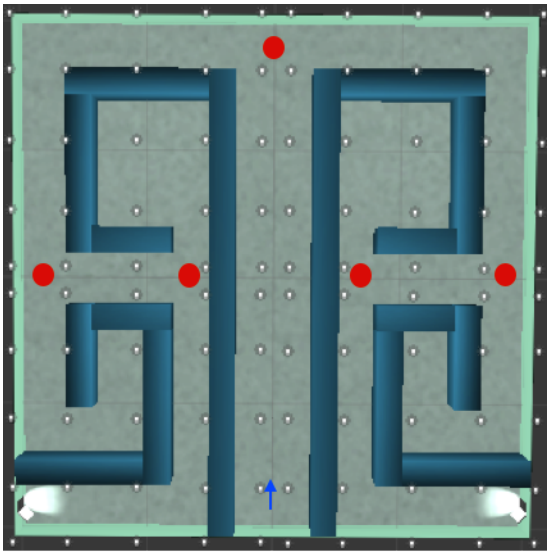


Figure 7.2: Maze environment for this experiment. Participants started in the position and orientation indicated by the blue arrow. Decision points are shown as red dots. A robot and an emergency exit sign with an arrow were at each decision point. One pointed to the path that lead to the exit on the left (shown in the diagram as an open door) and the other pointed to the exit to the right. Obstacles are shown in dark blue.

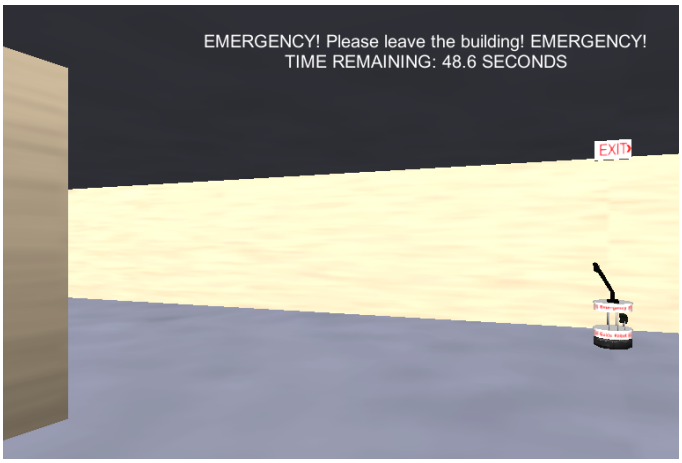
down from 60 seconds (see Figure 7.3). If a participant failed to find an exit in 60 seconds then the participant was informed that they had not survived the simulation. We validated this method for motivating participants in Section 6.3.

Three robots were tested: the Dynamic Sign robot, the Multi-Arm Gesture robot and the Humanoid. The Dynamic Sign platform was used because our prior research (Section 5.2) verified that participants would understand the information presented on the tablet. The multi-arm gesture robot was used because it also scored highly in previous tests. The Humanoid platform was included in order to test the difference, if any, between it and the Turtlebot-based Multi-Arm Gesture platform. The 3D models for each of the platforms were identical to those presented in Section 5.2.

After the interactive portion of the experiment, participants were asked four questions about their experience and then completed a short survey to gather demographic data. The four questions were:

1. Did you notice the robots doing anything to help you find the exit?
2. Did you notice the exit signs on the ceiling?
3. Did you trust the information provided by the robots?
4. Did you trust the information provided by the exit signs on the ceiling?

(a) Dynamic Sign Robot



(b) Multi-Arm Gesture Robot

(c) Humanoid Robot

Figure 7.3: Robots in the environment used for this experiment.

Each question offered yes and no options for a response and asked participants to explain themselves. The position and the time it took the participant to find the exit in the experiment was recorded. The environment was built in the Unity game engine and the participant interacted with the robot using a plugin in a web browser.

A total of 95 people participated in this experiment via Amazon's Mechanical Turk service. Five participants were unable to find the exit in the time provided and their results were excluded from analysis. The remaining 90 were evenly divided among the three robots.

7.2.2 Results

Overall, 61% of participants followed the robots instead of the exit signs ($p = 0.002$, Binomial test assuming 50% random chance of following either robots or signs for 90 samples). The difference in the following decision between the robots was not statistically significant at this sample size ($\chi^2(2, n = 90) = 0.341, p = 0.166$), but some trends can be gleaned from it. The Dynamic Sign robot had the highest following (77%), followed by the Multi-Arm Gesture (63%) and then the Humanoid (53%) (Figure 7.4). There was a strong correlation between noticing the exit sign and following the exit sign ($\phi(90) = 0.59$) and a weaker correlation between noticing the robots and following the robots ($\phi(90) = 0.39$). Figure 7.5 shows the results from participants who noticed the exit signs. Note that this sample size is relatively small, so it is hard to draw conclusions from the data, but most participants who noticed the exit signs chose to follow them.

Participants' explanations for their answers reinforced the conclusion that they did not notice the exit signs. Some representative comments were: "I didn't notice the exit signs on the ceiling. I would have followed them if I would have noticed," "I only saw one or two. These were not as helpful, since I was able to miss some," "I didn't think I had any other hint [besides the robots] on where to go," "In times of emergency, you have to make quick decisions so I chose to trust the robots." Other participants noticed the exit signs, but preferred to rely on the robots, saying, "I trusted them, but did not use them because I was [moving] too quickly to register them."

Some participants indicated that the exit signs had a greater chance of being correct by saying comments like, "[The robots] brought attention to the exit signs but appeared to be [pointing] in the wrong direction," "People wouldn't put up signs that pointed the wrong way," "I figured [the robots] to be more of a distraction and thought it would take too much time to figure out how they were trying to help me.," "It seemed like [the robots'] arm was moving. I [ignored] them though. The exit sign was easier to understand," "I decided to go by the sign on the wall because it was not moving

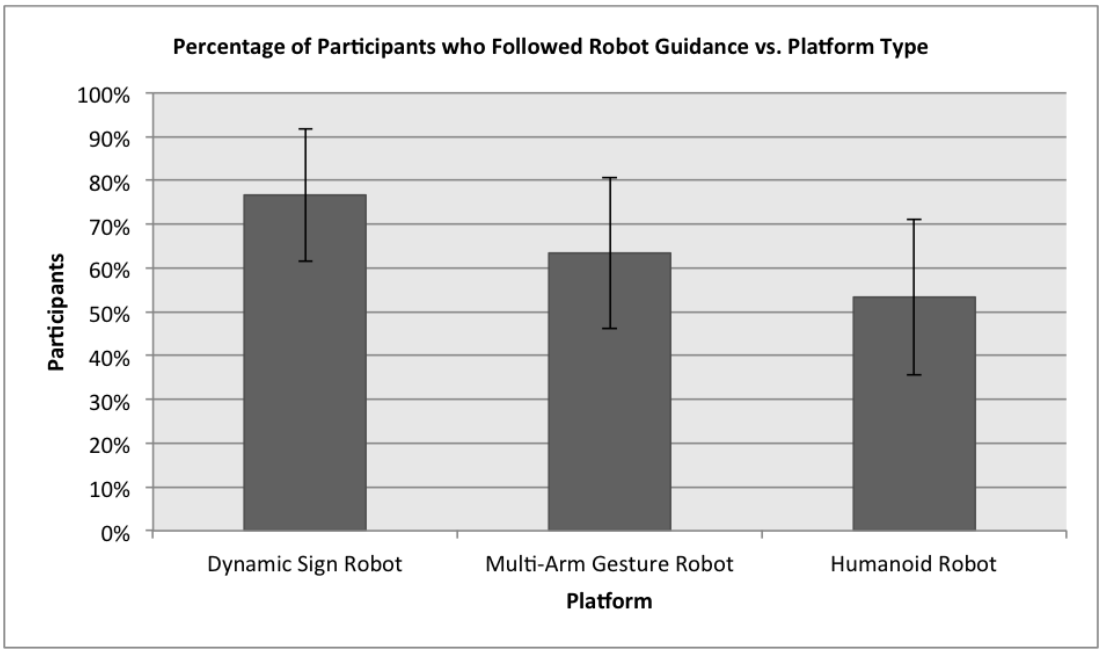


Figure 7.4: Percentage of participants who followed robot guidance broken down by robot type. Error bars represent 95% confidence intervals.

and seemed to be there longer.", "I didn't have time to figure out what [the robots] were trying to communicate." One participant indicated that he did not trust the robots because he had seen the film "I, Robot."

Other participants wrote that the robots, especially the Humanoid, looked like an authority figure: "[The robots] seemed to be with an 'authority' outfit, looked like policemen at first." Some did not understand that the dynamic sign robots were robots, and simply thought that it was a mobile sign: "They had the correct signs that were easily identifiable."

7.2.3 Discussion

Based on the comments and the correlations, we can conclude that participants generally followed the exit signs if they noticed the exit signs but were more likely to notice the robot. The robot was a sufficiently distracting object that most participants did not even notice the exit signs. We can thus conclude that robots are better at attracting attention during emergencies than standard emergency exit signs. These robots were also found to be sufficiently trustworthy to aid participants in finding an exit.

No significant difference was found between the three robots tested. In our later experiments, presented in the next two sections and in the next chapter, we use the Multi-Arm Gesture platform. This platform works well in both near and far conditions, as found in Chapter 5, and attracts

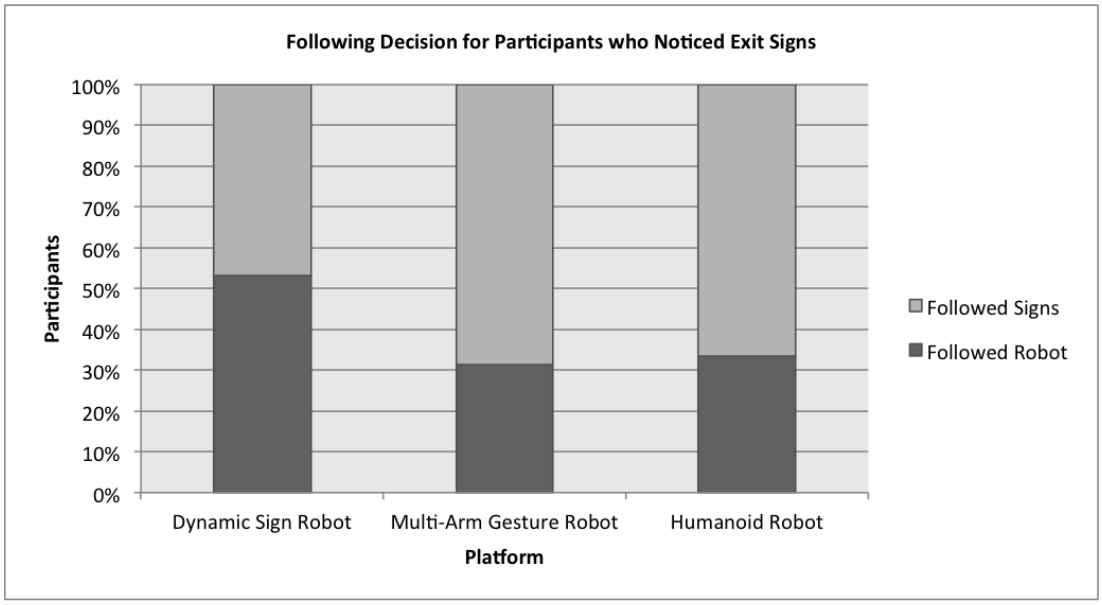


Figure 7.5: Results from participants who noticed exit signs only.

attention in emergency scenarios, as shown in this experiment. It has the additional benefit that it does not have to be facing participants to be understood, unlike the Dynamic Sign platform.

This experiment asked participants to navigate an unfamiliar building using both familiar (exit signs) and unfamiliar (guidance robots) technology. We found that most participants followed the robot because they did not notice the exit signs, and conclude that the robots attract attention in emergency situations. In the next section, we allow participants to gain some knowledge of the building before the emergency begins and allow participants to observe the robot's behavior before making their decision.

7.3 Virtual Office Evacuation Experiment

In Chapter 6, we asked participants to find an exit in an unfamiliar building with or without robot assistance and in the previous section we added emergency exit signs to that scenario. In this experiment, we ensured that participants are familiar with at least one path to an exit in our virtual environment by having participants start outdoors and proceed to a room inside. Participants followed guidance from a robot in this phase and could thus judge the robot's ability to be a good guide before the emergency started. When the emergency started, participants had the option of retracing their steps to the main entrance, following guidance from an emergency exit sign in the building, or following guidance from the robot.

Our previous results lead us to the following hypothesis: in a situation where participants are currently experiencing risk and have experienced a robot's behavior in a prior interaction, participants tend to follow guidance from an efficient robot but not follow guidance from a circuitous robot. As in Section 6.3, the robot performs efficient guidance by leading the participant straight to the destination and performs circuitous guidance by taking a less direct route.

7.3.1 Experimental Setup

To evaluate participant reaction to emergency guidance robots, we developed a three dimensional simulation of an office environment using the Unity game engine (Figure 7.6). The virtual office environment had a main entrance where the experiment began, several rooms to simulate offices and meeting rooms, and four emergency exits. Two emergency exits were marked with standard North American exit signs. The other two were unmarked. Additionally, the main entrance could be used as an exit. The robot guided participants through one half of the environment (containing one marked and one unmarked exit) in this experiment; however, participants were allowed to traverse the entire environment if they wished. The robot used in this experiment was a 3D model of a Turtlebot with signage identifying it as an emergency guide robot and two Pincher AX-12 arms to provide gestural guidance (this robot was introduced in Chapter 5 as the "Multi-Arm Gesture" platform).

The simulation began by introducing participants to the experiment and the robot. Participants were then asked to learn the movement controls of the simulation in a practice round. After the practice round, participants were asked to follow the robot to a meeting room where they were told they would receive further instructions. The robot's navigation behaviors during this phase are discussed below. After the participants reached the meeting room, the robot thanked them for following it and the participant was asked "Did the robot do a good job guiding you to the meeting room?" with space to explain their answers. Once the participants completed this short mid-experiment survey, they were told "Suddenly, you hear a fire alarm. You know that if you do not get out of the building QUICKLY you will not survive. You may choose ANY path you wish to get out of the building. Your payment is NOT based on any particular path or method." During this emergency phase, the robot provided guidance to the nearest unmarked exit. Participants could also choose to follow signs to a nearby emergency exit (approximately the same distance as the robot exit) or to retrace their steps to the main exit. As mentioned, above, other exits were available in the simulation, but participants were not expected to notice them as they would not have any

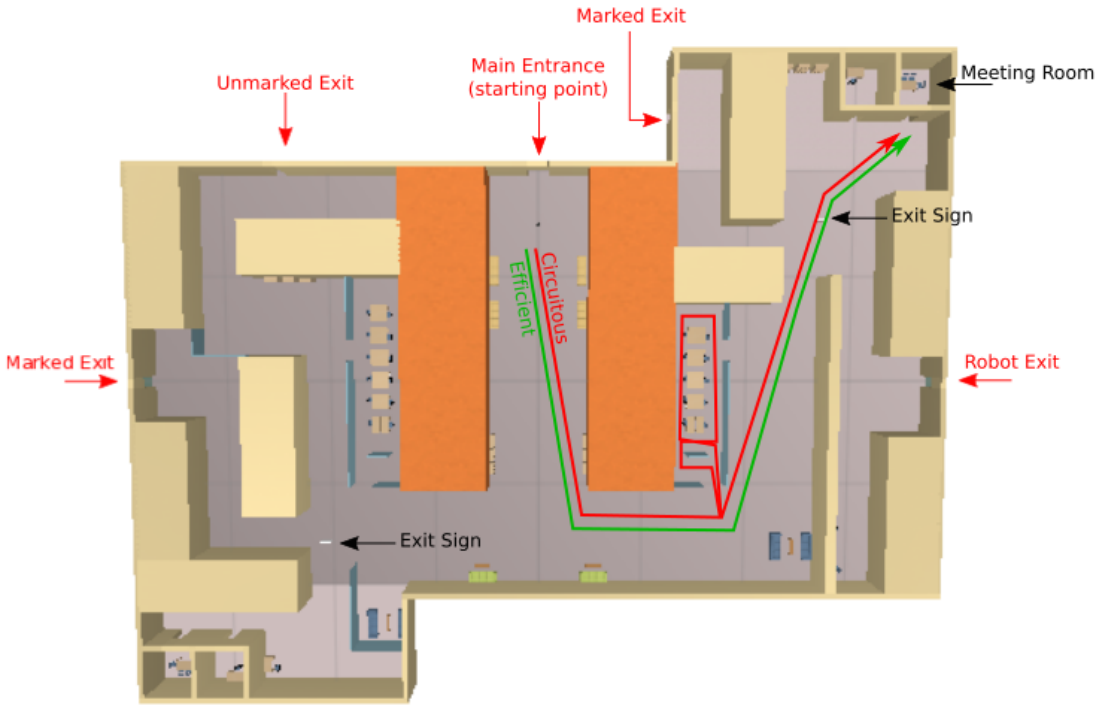


Figure 7.6: The virtual office environment used in the experiment. Efficient robot path (green) versus circuitous robot path (red) are shown.

reason to traverse that section of the environment. Participants were given 30 seconds to find an exit in the emergency phase (Figure 7.7). The time remaining was displayed on screen to a tenth of a second accuracy. This count down was shown in our previous research to have a significant effect in motivating participants to find an exit quickly (Chapter 6). The simulation ended when the participant found an exit or when the timer reached zero. After the simulation, the participants were informed if they had successfully exited or not. Finally, they were asked to complete a survey.

As in previous experiments (Chapter 6), the robot would either provide fast, efficient guidance to

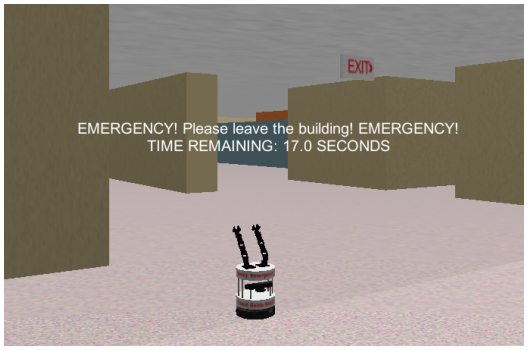


Figure 7.7: The robot providing guidance during the emergency phase. Participants had 30 seconds to exit. Note the clearly displayed emergency exit sign pointing to another exit.

the meeting room or take a circuitous route. In previous experiments, we showed that these behaviors can be used to bias most participants to trust (by using the efficient behavior) or not trust (by using the circuitous behavior) the robot later in the experiment. Efficient behavior consists of the robot guiding the participant directly to the meeting room without detours. Circuitous behavior consists of the robot guiding the participant through and around another room before taking the participant to the meeting room. Both behaviors can be seen in Figure 7.6. Each behavior was accomplished by having the robot follow waypoints in the simulation environment. At each waypoint, the robot stopped and used its arms to point to the next waypoint. The robot began moving towards the next waypoint when the participant approached it. The participant was not given any indication of the robot's behavior before the simulation started.

After the emergency phase of the simulation ended, participants were asked to complete a survey. All participants were asked a series of questions about how they found the exit (or attempted to), their motivation level during the emergency, and their opinion on the robot's ability to quickly find an exit. At the end of this survey, participants were asked to agree or disagree with the statement "I trusted the robot when I made my choice to follow or not follow the robot in the emergency." A third option was given for this question, allowing participants to indicate "Trust was not involved in my decision." Our previous work (Section 6.2) showed that participants are occasionally uncomfortable stating that they do not trust the robot, thus the inclusion of the third option. Our intent is to measure positive affirmations of trust in the robot. Therefore, we code all selections of the third option as not trusting the robot. Finally, participants were asked to answer demographic questions.

Similar to Chapter 6, our primary measure of trust is the participant's choice of exit where exiting through the door gestured at by the robot indicates they trusted the robot and exiting through any other door or not exiting at all indicates that they did not place trust in the robot in that situation. Additionally, we ask participants if they trusted the robot when they made their decision to follow or not follow it. In this experiment, that question is somewhat problematic: their decision to follow the robot or not is made over a continuous stretch of time from before they even see the robot until they have actually exited the building (or time has expired). Some participants may answer that question based on the decision they made while still in the meeting room while others may answer with their final decision. In prior work, we have found that answers to this question strongly correlate with decisions to use the robot when there is a single discrete decision (Chapter 6).

We deployed our simulation on the internet and solicited volunteers for our experiment via Amazon's Mechanical Turk service. Participants were paid $2.00 to complete this study. A total of 114 participants (56 in the efficient condition, 58 in the circuitous condition, average age of 32.4,

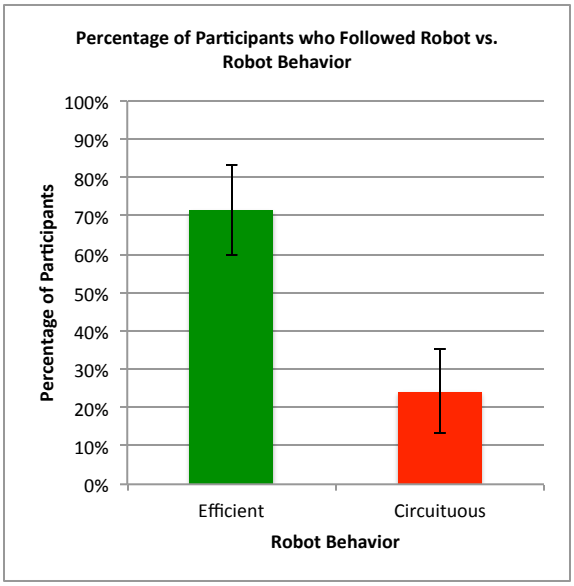


Figure 7.8: Results from the experiment. Error bars represent 95% confidence intervals.

38% female) were solicited on Amazon's Mechanical Turk service in a between-subjects experiment.

7.3.2 Results

A significant difference was found between the efficient (71% followed) and circuitous behavior (24% followed) ($\chi^2(1, 114) < 25.558$, $p < 0.001$), confirming our previous experiments (Section 6.3). For the efficient robots, 40 participants followed the robot, 15 followed the visible exit sign, and 1 retraced their steps to the main exit. For the circuitous robots, 14 participants followed the robot, 36 followed the visible emergency exit sign, 2 retraced their steps to the main exit, 3 found another exit out of the building, and 3 did not find an exit in time. Additionally, 55 of 56 (98%) participants indicated that the efficient robot did "a good job guiding" them to the meeting room (the one dissenter did not like that it stopped at each waypoint), compared with 21 of 58 (36%) participants who answered that the circuitous robot did a good job (only 8 of these, 14% of all participants in this category, followed it in the emergency). Many participants who indicated that the circuitous robot did a good job mentioned that it did seem to make an unnecessary detour, but that their overall experience was still positive. Only 12 of 58 participants (21%) indicated that they trusted the circuitous robot in the emergency phase, compared with 37 of 56 (66%) participants who indicated they trusted the efficient robot. These results are evidence that the use of circuitous robot behavior breaks trust and efficient robot behavior maintains trust in this experiment. All but one participant answered that they were motivated to find the exit in the emergency.

7.3.3 Discussion

The results from this experiment confirm results from Section 6.3: participants who previously experienced an efficient robot tend to follow it in an emergency but participants who previously experienced a circuitous robot will not follow it. This experiment confirms that this is true even when several other exits are made available to participants in an environment with which they have some familiarity. The experiment in Section 7.2 showed that robots attract more attention than exit signs, but this experiment shows that the robot's behavior can cause participants to look for other exits. Interestingly, most participants who did not follow the robot chose to follow the nearby exit sign instead of retracing their steps to the exit. This result differs from [7], which showed that people tend to exit through the front door of the building in emergencies.

Not all participants rated the circuitous robot as a bad guide, however most who rated it as a good guide still did not follow it to an exit. This indicates that they did not consider the robot's previous error to be important when it guided them to the meeting room, but that the error was bad enough to convince them to find their own way out in an emergency.

As in previous work (Chapter 6), we believe that participants took this experiment seriously. All but one participant responded that they were motivated to find an exit quickly in the emergency and only three participants did not manage to find an exit in time. Nevertheless, participant interaction with the emergency scenario was mediated by a computer, and thus participants probably did not feel that they needed to evacuate as urgently as they would in a physical scenario. The next section presents a similar experiment performed in the physical domain.

7.4 Physical Office Evacuation Experiment

To create a high-risk situation, we conducted a real-world emergency evacuation scenario using fire alarms and artificial smoke to add urgency. This was conducted in a manner similar to the previous experiment (Section 7.3), but the smoke and alarms provided increased motivation for participants to find an exit. Participants were not informed that an emergency would take place prior to the experiment.

This experiment had the same hypothesis as the previous experiment: in a situation where participants are currently experiencing risk and have experienced a robot's behavior in a prior interaction, participants will tend to follow guidance from an efficient robot but not follow guidance from a circuitous robot.

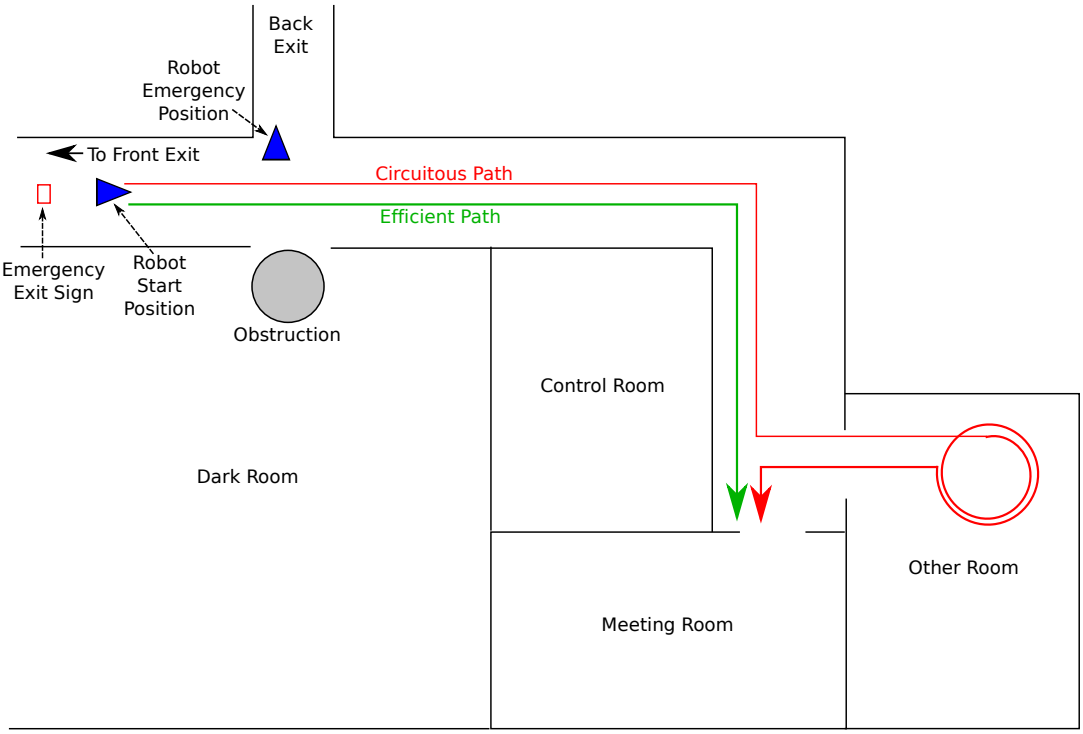


Figure 7.9: Layout of experiment area showing efficient and circuitous paths.

7.4.1 Experimental Setup

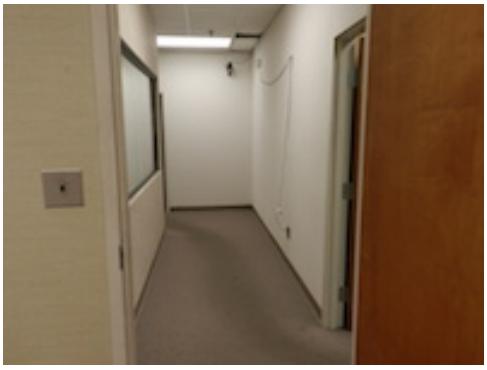
This experiment took place in the office area of a storage building on the Georgia Tech campus. The building was otherwise unoccupied during experiments. The office area contained a hallway and several rooms (Figures 7.9 and 7.10). The room at the end of the hallway was designated the meeting room and the room next to it was designated the other room, only used in the circuitous behavior condition. The back exit used for this experiment actually lead to a large storage area, but this was obscured using a curtain. Participants could see light through the curtain, but could not see the size of the room. This was intended to make this doorway into a plausible path to an exit, but not a definite exit to the outdoors. A standard green emergency exit sign hung in the hallway indicating that participants should exit through the main entrance in the event of an emergency. A room in the middle of the building was designated as the control room. An experimenter stayed in that room controlling the robot through an RF link. The experimenter could view the entire experiment area from five cameras placed throughout the building but could not be seen by participants.

The emergency guide robot (Figure 7.11) used a Pioneer P3-AT as a base. The base had white LED strip lights along all sides to illuminate the ground around it. A platform was built on top of this base to house a laptop computer and support a lighted cylindrical sign 24.5 cm tall and 47 cm

(a) Exterior of building

(b) Main hallway

(c) Interior of meeting room



(d) View of hallway from meeting room door

Figure 7.10: Pictures of the experiment site

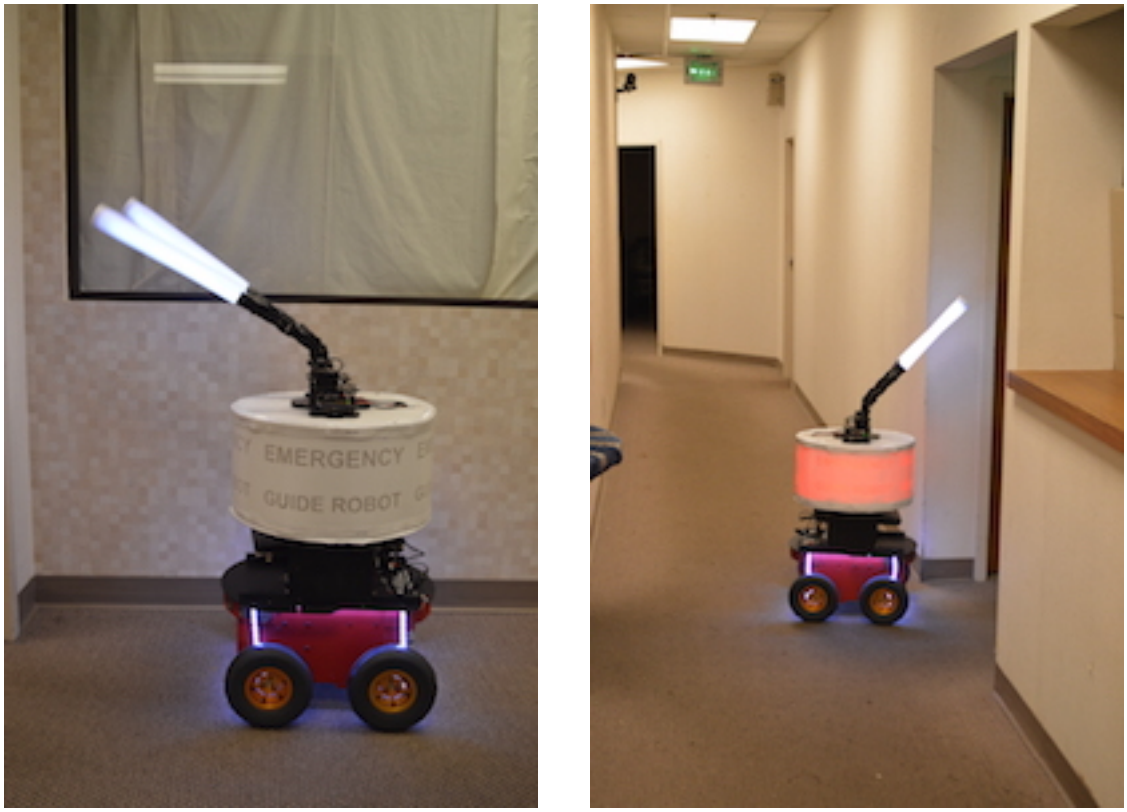


Figure 7.11: Robot during non-emergency phase of the experiment pointing to meeting room door (left) and robot during emergency pointing to back exit (right). Note that the sign is lit in the right picture. A standard emergency exit sign is visible behind the robot in the emergency.

in diameter. The words "EMERGENCY GUIDE ROBOT" in 3 cm tall letters were backlit by red LEDs. These LEDs were off during the non-emergency phase of the experiment but turned on during the emergency. Two PhantomX AX-12 Pincher arms were mounted to the top of the sign. Only the first three joints (the two shoulder servos and the elbow servo) on each arm were present. On top of each arm was a cylinder of foam lit with white LEDs. The arms, including foam, were 68 cm long. While the robot was moving the arms stayed straight up. The arms pointed straight ahead and oscillated by 20 degrees up and down to indicate that a participant should proceed in the direction the robot is facing (either into the meeting room or to the back exit). The robot measured 68 cm from ground to the top of the sign and 136 cm tall with arms fully extended up. This robot was modeled after the Multi-Arm Gesture platform in Chapter 5 so that it could provide understandable guidance to the participant from a distance or from a nearby position. For participant safety, the robot was teleoperated for the entire experiment.

Participants began the experiment by reading and signing a consent form. They then completed Survey 1, which asked them to agree or disagree with ten statements about robots (Table 7.2) and

Table 7.2: List of statements about robots. Participants agreed or disagreed with each before, during, and after the experiment.

Robots are useless
Robots are trustworthy
Robots are pretty
Robots are deceptive
Robots are too fast
Robots are dangerous
Robots are kind
Robots are creepy
Robots are helpful
Robots are safe

Table 7.3: Statements about feelings given before (Survey 1), during (Survey 2), and after the experiment (Survey 3). Participants rated their agreement with the statement on a 7 point Likert scale.

Survey 1	Survey 2	Survey 3
I feel comfortable right now.	I felt comfortable while moving through the building.	I felt comfortable during the emergency.
I feel relaxed right now.	I felt relaxed while moving through the building.	I felt relaxed during the emergency.
I feel safe right now.	I felt safe while moving through the building.	I felt safe during the emergency.
I feel confused right now.	I felt confused while moving through the building.	I felt confused during the emergency.
I feel scared right now.	I felt scared while moving through the building.	I felt scared during the emergency.

asked them to rate their current feelings on a seven point Likert scale (Table 7.3). Five feelings were included on the survey: comfortable, safe, relaxed, confused, and scared. Upon completing the survey, participants were given a device that would produce a loud sound in case they needed to abort the experiment for any reason, introduced to the robot and instructed to follow it to a meeting room. They were told that the robot would inform them when they had reached the meeting room by pointing with its arms (this gesture was previously evaluated in Chapter 5). Participants were told that written instructions were in the meeting room.

During this non-emergency guidance phase, participants followed the robot to the meeting room. The robot was remote controlled in order to ensure participant safety. The robot operator was in a room hidden from participants. The robot performed either efficient or circuitous guidance along the way (see Figure 7.9 for the exact guidance performed in this experiment). Participants were randomly assigned to each condition.

After arriving at the meeting room, participants followed written instructions labeled as "Meeting Room Instructions" that were posted in two locations on the wall as well as on a table:

Meeting Room Instructions

1. Sit down at the table.
2. Complete the survey on the clipboard.
3. Close the door to this room.
4. Read the article on the table. Mark important sections of the article and make any notes necessary on the page. You will be asked questions about this document after the experiment is completed.

This survey (Survey 2) first asked participants "Did the robot do a good job guiding you to the meeting room?" and why or why not. Then the ten robot statements from Survey 1 were asked. Finally, participants were asked how they felt (using the same five feelings as before) while moving through the building to the meeting room. They were then asked to read an article from the IEEE Spectrum magazine about indoor navigation technologies [66]. It did not mention emergency guide robots and was chosen because it was somewhat related to the non-emergency phase of the experiment, but was unlikely to bias participants towards or against the robot.

A timer on a smoke generator was triggered when the participant closed the meeting room door. This typically occurred after the participant finished the survey and before they started the article, but some participants closed the door early. Other participants never closed the door and an experimenter closed it when it was clear from camera footage that participants had completed the survey. The timer counted down for three minutes and then released artificial smoke into the hallway outside the meeting room door for twelve seconds. In tests, this was found to be sufficient to give a smoky appearance to the hallway outside the meeting room but not enough to impair participant vision such that they would be likely to injure themselves (see Figure 7.12 for example). The robot was placed in the hallway along the path to the main entrance and pointed down another hallway to what appeared to be another exit (Figure 7.9).

Artificial smoke was created using a Bullex SG6000 smoke generator. The artificial smoke is non-toxic and non-carcinogenic. Two First Alert smoke detectors were used in the experiment. One was placed on the hallway side of the doorframe of the meeting room door. The other was placed in the other room on the wall in case the first did not sound. The detectors alternated between producing a buzzing noise and the words "Evacuate! Smoke! Evacuate!" when they detected smoke. The alarm could easily be heard in the meeting room with the door closed.

The emergency phase of the experiment began when artificial smoke triggered a smoke detector.

Figure 7.12: Example of smoke-filled hallway after smoke detector is triggered.

This took approximately thirty to ninety seconds after the smoke stopped. Participants exited the room, proceeded down the hallway to the first corner, and then observed the robot. They then decided to either follow its guidance or proceed to the main entrance.

An experimenter was waiting at the entrance and another was waiting at the back exit, where the robot was pointing, during the emergency phase. When the participant had clearly made their choice by walking down the hallway to the main entrance or down the hallway to the back exit, the closest experimenter stopped him or her and informed him or her that the emergency was a part of the experiment. The participant was then given a third survey, where they were asked about the realism of the emergency, the method they used to find an exit, whether their decision to use the robot indicated that they trusted it, the same ten statements as before, the five questions on feelings, and demographic information.

Aside from three experimenters and one participant, no one else was in the building at the time of the experiment. The study was performed in a double-blind manner in which neither the experimenters that interacted with the participants nor the participants themselves knew what type of behavior the robot employed. Participants were not warned that an emergency would occur. This experiment was approved by the university's IRB and was conducted under the guidance of the Georgia Tech Fire Marshal.

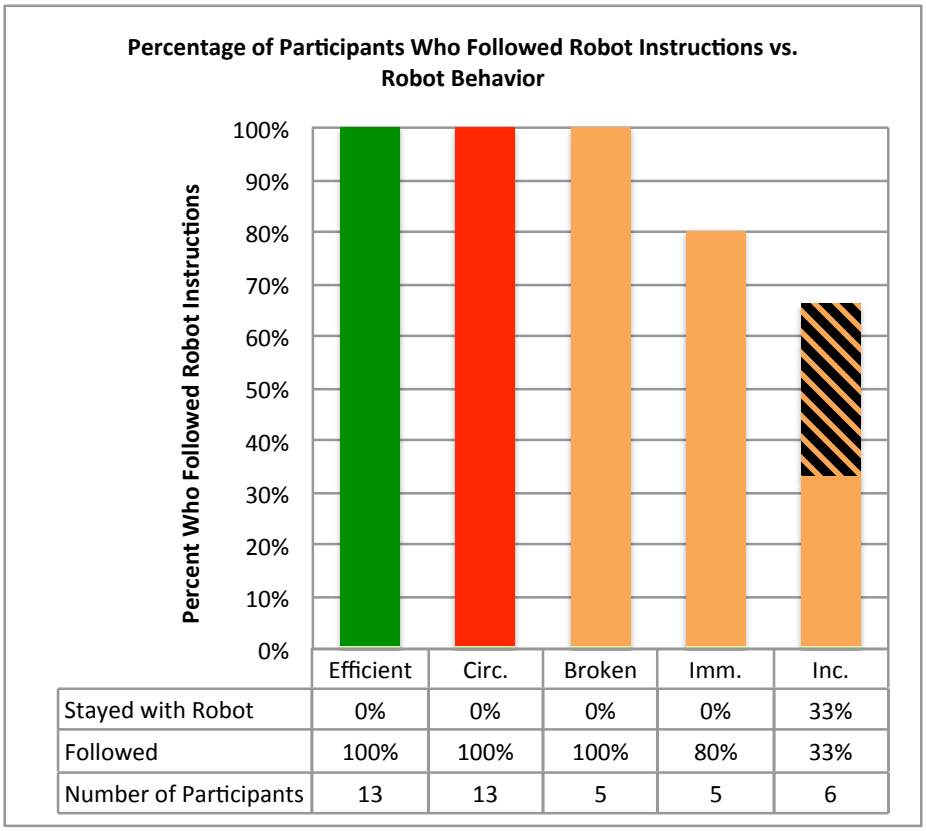
Participants were warned before signing up for the experiment and in the consent form that they should not participate in this experiment if they have known heart conditions, asthma, other respiratory conditions, Post-Traumatic Stress Disorder (PTSD), anxiety disorders, or claustrophobia.

They were not told why. These exclusion criteria were put in place because the artificial smoke can irritate people with respiratory disorders and because the emergency scenario could negatively affect participants with heart conditions or psychological disorders. Participants were also required to be between the ages of 18 and 50 (for health reasons) and capable of simple physical activity, such as walking around the building. The exclusion criteria was intentionally designed to be restrictive to ensure participant safety to the extent possible.

Participants were recruited via emails to students at the university. Thirty participants were recruited for this study but four were not included in the results because they did not complete the experiment. Two participants did not leave the meeting room after the alarm sounded and had to be retrieved by experimenters. One participant activated the abort device after walking through the smoke and was intercepted by an experimenter before completing the experiment. In one trial, neither alarm sounded after the smoke filled the hallway, so the experiment was aborted. Of the 26 remaining participants (31% female, average age of 22.5), 13 were in each condition. All but three participants stated they were students.

7.4.2 Results

The results from this experiment were surprising: all 26 participants followed the robot's instructions to proceed to the back exit in the emergency (Figure 7.13). This result is significantly greater than either the circuitous robot ($\chi^2(84,1) = 41.421, p < 0.001$) or the efficient robot ($\chi^2(82,1) = 9.229, p = 0.002$) from the virtual experiment presented in Section 7.3. Eighty-one percent of participants indicated that their decision to follow the robot meant they trusted the robot. The remaining five individuals (three in the efficient condition, two in the circuitous condition) stated that trust was not involved. These five people justified this with a variety of different reasons. One participant in the circuitous condition stated that they did not believe that the emergency was real. One in the efficient condition felt that they had no choice in the emergency. Another in the efficient condition noted that following the robot was the logical choice. One participant (also in the efficient condition) indicated that the robot was designed to help (and thus it was not the robot that was being trusted) and the last (in the circuitous condition) believed that trust was not involved in this interaction because they would not necessarily trust the robot in every emergency. Eighty-five percent of participants indicated that they would follow the robot in a future emergency. Only three participants noticed the emergency exit sign behind the robot and none expressed an interest in following it.



	Efficient	Circ.	Broken	Imm.	Inc.
Stayed with Robot	0%	0%	0%	0%	33%
Followed	100%	100%	100%	80%	33%
Number of Participants	13	13	5	5	6

Figure 7.13: Results from the main study (green and red bars) and exploratory studies (orange bars) discussed in the next section.

Results from the second survey found that just four of the thirteen participants with the circuitous robot reported that it was a bad guide. Three other participants indicated that it was a good guide in general, but that it made a mistake by going into the other room. The remaining six participants gave varying reasons for why they thought the robot was a good guide, including that it moved smoothly and pointed to the right room in the end. It is worth noting that in Section 7.3, we found that many participants marked that the robot was a good guide in the non-emergency phase of the experiment, but were still biased against following it in the emergency. This result inspired one of the exploratory studies presented in the next section.

There are confounding factors that could serve as alternative explanations for the results and explain why participants behaved differently in this experiment than in previous virtual emergency experiments such as in Section 6.3 and Section 7.3. Lack of believability during the experiment is one confounding factor. Participants may not have believed the simulated emergency was real and based their decision and survey responses as such. It is difficult to measure the realism of the experiment because participants may not want to admit that they were deceived (social desirability bias). We attempted to evaluate the experiment's believability by asking participants to complete a survey about their current feelings before and after the experiment. The change in these results can be seen in Figure 7.14. All of the survey questions were on a 7-point Likert scale. Participants generally reported being comfortable, relaxed and safe before the experiment began (median of 6 for each). Some participants reported being confused (median of 3) and almost none reported being scared (median of 1) in the beginning. There was very little change (median changed less than or equal to 1 on each question) in the second survey. Participant answers in the third survey showed a marked change in comfort, relaxation, and safety level, (median of 5, 4, and 5, respectively), and increase in confusion (median of 4.5) with a similar increase for the scared scale (median of 2.5). Fifty-four percent of participants gave an increased confusion score between the pre and post surveys with 27% (seven participants) increasing that score by 3 or more. Additionally, 62% of participants (mainly the same participants) increased their response to the scared question with 15% increasing their rating by 3 or more. Wilcoxon Signed-Ranks Tests indicate that these results were significant: Comfortable $Z = 12, p < 0.001$, Relaxed $Z = 22, p = 0.003$, Safe $Z = 26, p < 0.001$, Confused $Z = 35.5, p = 0.023$, Scared $Z = 4.5, p < 0.001$

Despite this decrease in positive feelings and increase in negative feelings, most participants (58%) rated the realism of the emergency as low (a 1 or 2). Thirty-eight percent of participants rated it as moderate (3, 4 or 5) and only one participant rated it as high (a 6). The one participant who aborted the experiment (not included in the results above due to not completing the experiment)

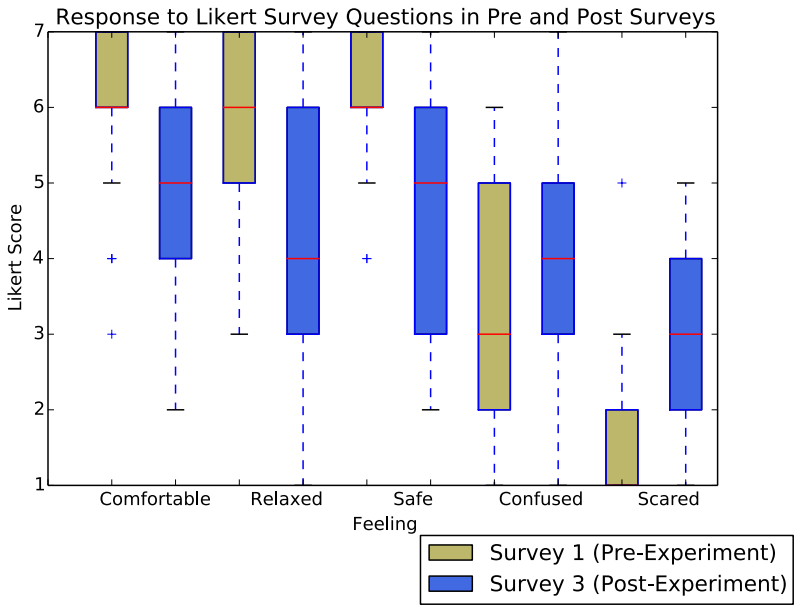


Figure 7.14: Change in participant responses to questions about their feelings from before the experiment (gold) to the emergency (blue).

after seeing the smoke rated it a 6 out of 7. After reviewing video recordings of the experiment, we observed that 42% of participants had a clear physical response (either leaning away from the smoke or stepping back from the door in surprise) when opening the door to a smoke-filled hallway. This leads us to believe that many participants were likely exhibiting post-hoc rationalization on the survey: when they took the survey they knew that the experiment was not real, so they responded accordingly. The percentage of participants who agreed with the statements about robots in the three surveys can be seen in Table 7.4. The differences between surveys were not significant at a $p < 0.05$ level even though an increase was seen in the statements about robot trustworthiness and safety.

7.4.3 Exploratory Studies: How to Bias Against Following the Robot

The results above are promising from one perspective: clearly the robot is considered to be trustworthy in an emergency. Yet, it is concerning that participants are so willing to follow a robot in a potentially dangerous situation even when it has recently made mistakes. The observed phenomena could be characterized as an example of overtrust [44]. Overtrust occurs when people accept too much risk believing that the trusted entity will mitigate this risk. This raises the important question: how defective must a robot be before participants will stop following its instructions?

Table 7.4: Percent agreement with statements about robots before, during, and after the experiment.

Statement	Survey 1	Survey 2	Survey 3
Useless	4%	4%	4%
Trustworthy	60%	72%	85%
Pretty	40%	48%	38%
Deceptive	32%	20%	8%
Too Fast	16%	4%	8%
Dangerous	32%	32%	31%
Kind	40%	44%	54%
Creepy	16%	16%	12%
Helpful	96%	96%	96%
Safe	68%	88%	85%

Following our main study, we conducted three small studies to determine if additional behaviors from the robot either before or during the emergency would convince participants not to follow its instructions in the emergency. The first exploratory study, labeled Broken Robot, tested a new behavior during the non-emergency phase of the experiment. The second, Immobilized Robot, evaluated a behavior that spanned the entire study. The final study, Incorrect Guidance, determined the effect of odd robot behavior during the emergency phase of the experiment. A total of 19 participants were recruited for the three studies but three did not complete the experiment. One because the alarm failed to sound, one because the participant left the meeting room before the emergency started and one because the participant did not leave the meeting room after the alarm sounded. The 16 remaining participants (38% female, average age of 20.9 years old) were divided into the three new conditions.

7.4.3.1 Broken Robot

We believed that the robot's behavior during the non-emergency phase of the experiment would influence the decision-making of the participant during the emergency. Given that 54% of the participants did not realize that the circuitous robot had done anything wrong, we designed a robot behavior that would be explicitly identified as a bad guide. As with the main experiment, this experiment began by guiding participants down the hallway. When it reached the first corner, the robot spun in place three times and pointed at the corner itself (Figure 7.15). No discernible guidance information was provided by the robot to participants. An experimenter then approached the participant and said, "Well, I think the robot is broken again. Please go into that room [accompanied with gestures to the meeting room] and follow the instructions. I'm sorry about that." The experiment then proceeded as in the main experiment with the robot moving to the Robot Emergency Position

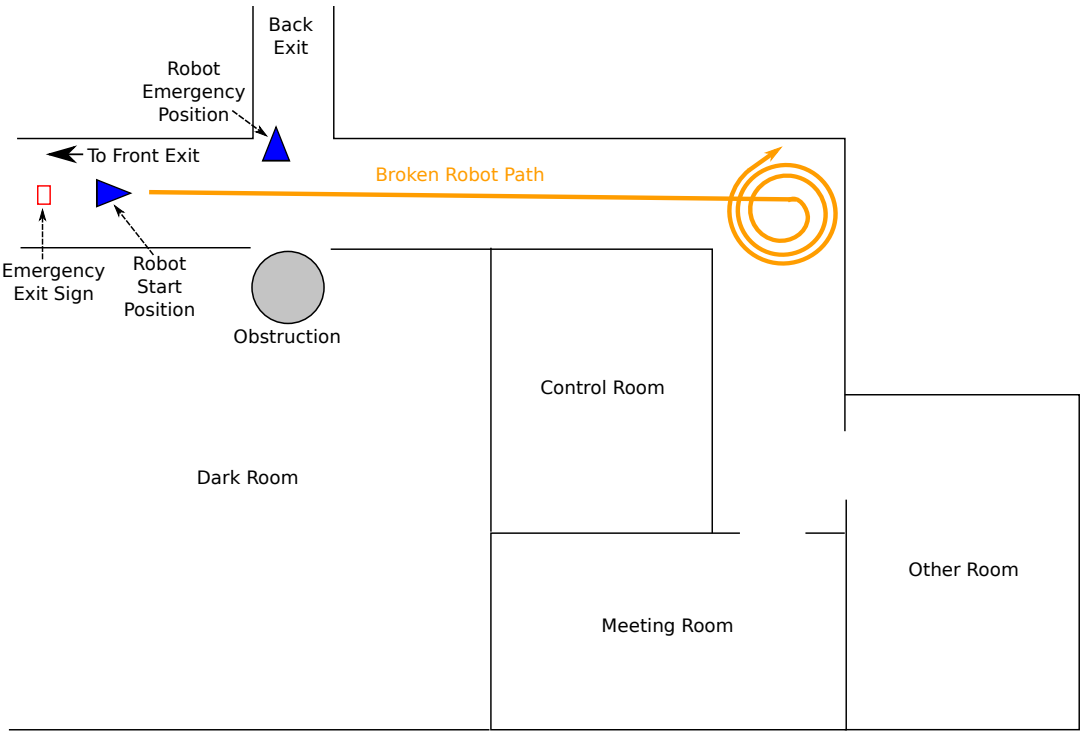


Figure 7.15: Robot path during the broken robot study. The robot spins in place at the first corner and the participant is then informed that the robot is broken. The robot is placed in the same emergency position as in the main experiment.

after the participant closed the meeting room door. Five participants took part in this study.

During the emergency, despite the robot's breakdown in the non-emergency phase of the experiment, all five participants followed the robot's guidance by exiting through the back exit. All five indicated that their decision meant that they trusted the robot and all five indicated that they would follow it in a future emergency. Four of the five participants indicated that the robot was not a good guide in the non-emergency phase of the experiment. The only one who indicated that it was a good guide did not hear the speech from the experimenter and thus did not experience the entire robot condition. The participant saw the robot spin in circles and then found the meeting room without any help. He considered that the robot had done a good job because he was able to find the meeting room quickly. Despite the higher percentage of participants who rated the robot as a bad guide in the non-emergency phase of the experiment, this condition produced the same results as in the circuitous condition.

Participants rated the emergency with a median of 3 out of 7 on the realism scale. Participants rated their feelings in the emergency scenario with a median of 5 for comfort, 5 for relaxation, 6 for safety, 4 for confusion and 4 for scared.

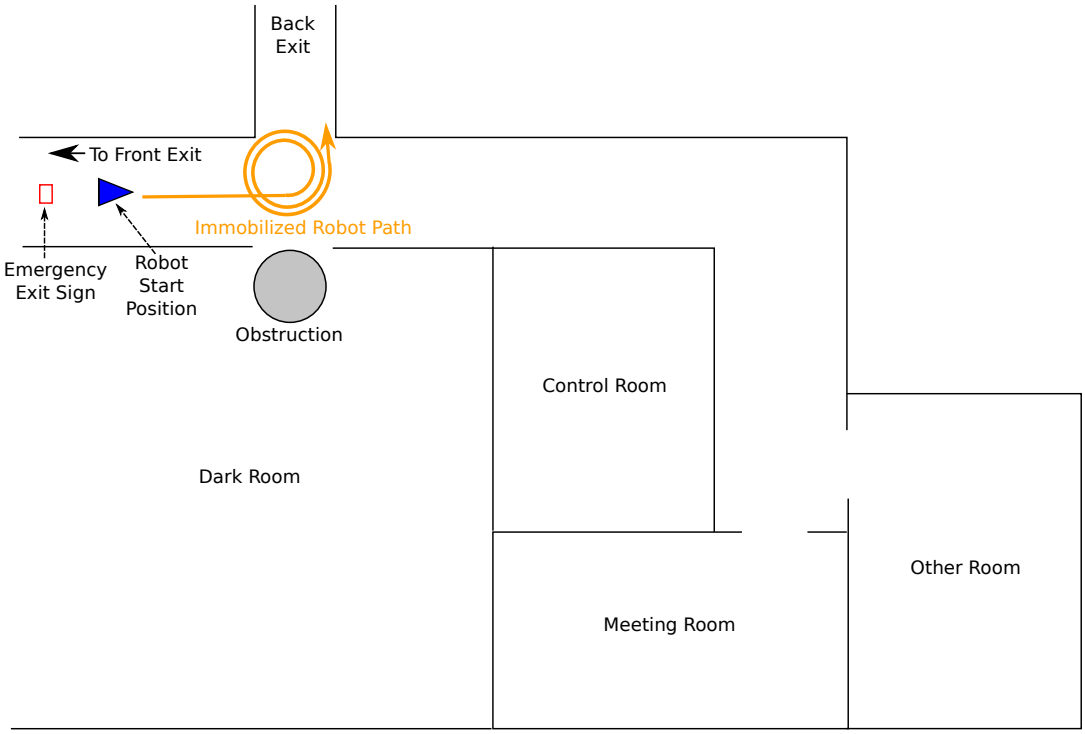


Figure 7.16: Robot path during the immobilized robot study. The robot spun in place at its normal emergency position and then remained there, pointing towards the back exit, for the rest of the experiment. After watching the robot, the participant was told that it was broken.

7.4.3.2 Immobilized Robot

In the immobilized robot study, we attempted to convince participants that the robot was still malfunctioning during the emergency by having it behave poorly throughout the experiment.

At the start of the experiment, the robot moved a short distance forward, but then, upon reaching the intersection of the hallways (Robot Emergency Position in Figure 7.11) it spun in place three times and then pointed to the back exit. At this point, an experimenter informed the participant that the robot was broken with a similar speech as in the broken robot study. The robot did not move and continued gesturing towards the back exit for the remainder of the experiment. The robot's lights were not turned on. From the perspective of an evacuating participant, the robot did not appear to have moved or changed behavior from when they were told it was broken in the non-emergency phase of the experiment. Five participants took part in this study.

Four of the five participants followed the robot in the emergency. The one participant who did not follow the robot noticed the exit sign and chose to follow it instead. Three of the four participants who followed the robot's guidance indicated that they trusted it (the remaining said that this was the first exit available and thus trust was not involved). Two said they would follow it again in the

Figure 7.17: Robot providing incorrect guidance condition by pointing to a dark, blocked room in the emergency.

future. All five rated the robot as a bad guide in the non-emergency phase of the experiment.

Participants rated the emergency with a median of 1.5 out of 7 on the realism scale. Participants rated their feelings in the emergency scenario with a median of 3 for comfort, 3 for relaxation, 5 for safety, 6 for confusion and 4 for scared.

7.4.3.3 Incorrect Guidance

Building on the results in the immobilized robot study, we tried a third robot behavior to convince participants not to follow its guidance in an emergency. In this study, the robot performed the same as in the broken robot condition, with accompanying experimenter speech, in the non-emergency phase of the experiment. During the emergency, the robot was stationed across the hall from its normal emergency position and instructed participants to enter a dark room in front of it (see Figures 7.9 and 7.17). The doorway to the room was blocked in all conditions with a piece of furniture (initially a couch then a table when the couch became unavailable) that left a small amount of room on either side for a participant to squeeze through to enter the room. There was no indication of an exit from the participant's vantage point. All lights inside the room were turned off. Six participants took part in this condition.

Two of six participants followed the robot's guidance and squeezed past the couch into the dark room. An additional two participants stood with the robot and did not move to find any exit on their own during the emergency. Experimenters retrieved them after it became clear that they would not leave the robot. The remaining two participants proceeded to the front exit of the building. The two participants who entered the dark room indicated that this meant they trusted the robot, although one said that he would not follow it again because it had failed twice. The two who stayed

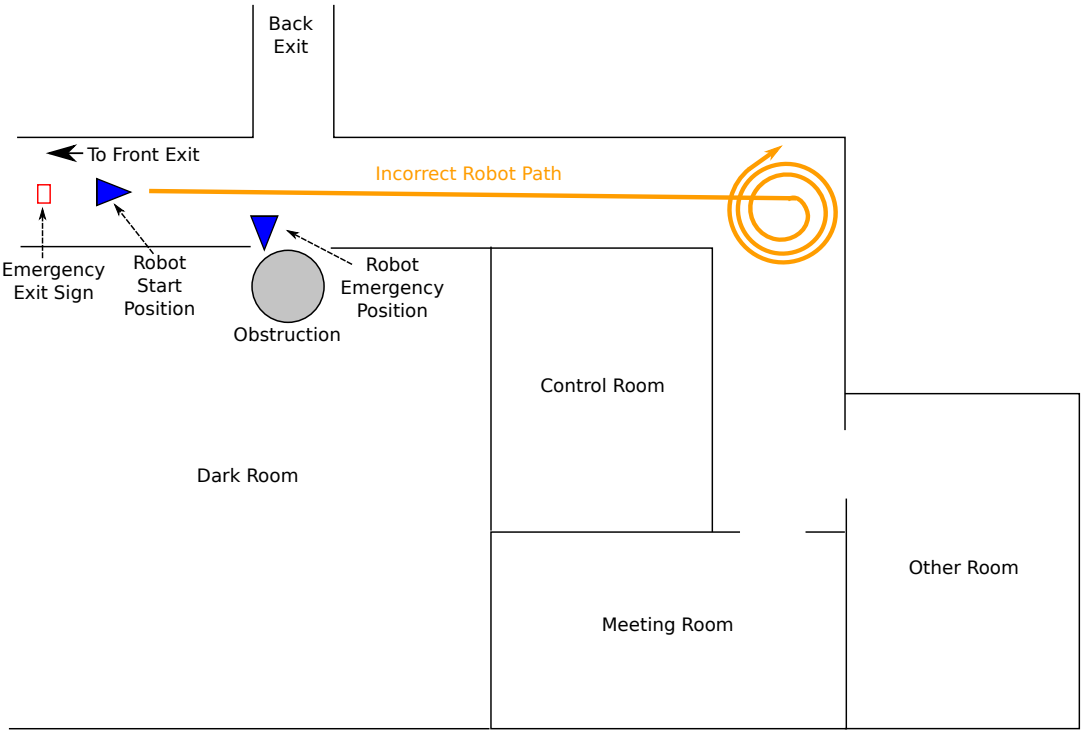


Figure 7.18: In the Incorrect Guidance study, the robot performed the same as in the Broken Robot study for the first phase, but then pointed to the dark room in the emergency phase.

with the robot indicated that they did not trust the robot and the two who proceeded to the front exit selected that trust was not involved in their decision. None of those four indicated that they would follow the robot in a future interaction. All six participants wrote that the robot was a bad guide in the non-emergency phase of the experiment.

Participants rated the emergency with a median of 1.5 out of 7 on the realism scale. Participants rated their feelings in the emergency scenario with a median of 4 for comfort, 4 for relaxation, 5 for safety, 5.5 for confusion and 3 for scared.

7.4.4 Discussion

Our results show that none of the robot behaviors performed solely in the non-emergency phase of the experiment had an effect on decisions made by participants during the emergency. These results conflict with our hypothesis and offer evidence that errors during prior interactions have little effect on a person's later decision to follow the robot's guidance. These results appear to disagree with the work of others examining operator-robot interaction in low-risk situations [23] and emergency guidance studies in virtual simulation environments (Chapter 6 and Section 7.3). A similar result was found in [65].

Our results show that participants have a strong tendency to follow a robot's guidance regardless of its prior behavior. To better understand participants' reasoning, we examined their survey responses. Of the 42 participants included in all of our studies, 32 (76%) reported not noticing the exit sign behind the robot's emergency position. Upon turning the corner from the smoke filled hallway on their way out, participants' eyes were drawn to the large, well-lit, waving robot in the middle of their path. Couple the visual attraction of the robot with the increased confusion reported on the surveys and it is no surprise that participants latched onto the first and most obvious form of guidance that they observed.

These results are in contrast to our previous results from a virtual simulation of an emergency that found participants did not follow a previously circuitous robot. In the high-risk scenario investigated here, participants observed what appeared to be smoke and had to make fast decisions. These types of situations tend to invoke fight-or-flight responses. Hence, the decisions made under these conditions may be qualitatively different from those made during a virtual emergency. Although the virtual emergency was also under time pressure, participants were not in real danger and thus were able to be more deliberative in their decision-making. They were likely conscious of the fact that they were in no real danger and so they could take their time to make the best choice possible.

Before this conjecture can be affirmed, several alternative explanations for the results must be ruled out. One alternative explanation is that the age of the subject population caused the observed results. Participants in this study were mostly university students and therefore younger and possibly more accepting of new technology than a more diverse population. Still, even if our findings are only true in relation to a narrow population, they show a potentially dangerous level of overtrust. On average, participants in the physical study were almost 10 years younger than their virtual counterparts. We might, therefore, expect that these participants would be more accepting of new technology and thus more likely to trust new technology. Both sets of participants were asked to rate their agreement with the statement "I am comfortable with using new technology" on a 7 point Likert scale. Table 7.5 shows that their responses were not very different ($H(2) = 0.003, p = 0.958$). A large majority of each set of participants (89% in virtual, 88% in physical) rated their comfort with technology as a 6 or 7. This should come as no surprise as participants in the virtual experiment had to be sufficiently comfortable with technology to perform tasks on the web-based Mechanical Turk. Therefore, we conjecture that differing attitudes on technology were not responsible for the difference in results between the two studies.

The realism of the scenario is addressed in detail above, but still presents an alternative explanation. Perhaps participants did not believe that they were in any danger and followed the robot

Table 7.5: Participant responses when asked to rate their comfort with technology on a 7 point Likert scale

Rating	Percentage of Participants	
	Virtual	Physical
<5	3%	0%
5	8%	12%
6	22%	27%
7	68%	62%

for other reasons. Their increased confusion scores and reactions to the smoke indicate that at least some of the participants were reacting as if this was a real emergency. Given that every participant in the main study followed the robot, regardless of their rating of emergency realism, we believe that the realism of the scenario had little or no effect on their response. Additionally, many participants wrote that they followed the robot specifically because it stated it was an emergency guide robot on its sign. They believed that it had been programmed to help in this emergency. This is concerning because participants seem willing to believe in the stated purpose of the robot even after they have been shown that the robot makes mistakes during a related task. One of the two participants who followed the robot's guidance into the dark room even thought that the robot was trying to guide him to a safe place after he was told by the experimenter that the exit was in another direction. Most participants in the physical experiment reported that they did not believe the emergency was real, but if the same question had been asked in the virtual experiment we would expect none of them to believe that the emergency was actually real. The emergency was contained to their computer and thus could not affect them in any way. Interestingly, significantly more participants in the virtual experiment reported that they were motivated in the emergency phase of the experiment than in the physical experiment ($\chi^2(140, 1) = 26.658, p < 0.001$).

It is worth mentioning that many people in real-life fire drills and fire emergencies do not believe that they are in real danger (see [27] for an example using the 1993 World Trade Center bombing). Some participants wrote on their surveys that the fire alarm used in this experiment sounded fake, even though it was an off-the-shelf First Alert smoke detector. Others stated that the smoke seemed fake, even though the same artificial smoke is used to train firefighters. It is likely that participants would respond the same when encountering real smoke.

Perhaps participants only followed the robot because they felt that they should do so in order to complete the experiment. One participant of the 42 tested wrote that he followed the robot only because he was told to in the non-emergency phase of the experiment. Each of the conditions in the exploratory studies attempted to break or realign participant beliefs by having the experimenter

interrupt the robot and lead the participant himself. In the broken and immobilized robot case, nine of ten participants still followed the robot in the emergency.

Another alternative explanation is that the building layout was sufficiently simple that participants believed that they had ample time to explore where the robot was pointing and still find their way out without being harmed. This is possible, but participants did not express a desire to explore any other rooms or hallways in the building, just the one pointed to by the robot. Some participants looked into the other room on their way out, but none spent time exploring it. No participant tried to open either of the closed doors on their way out and, except in the incorrect guidance case, no participant tried to enter either of the rooms blocked by furniture. Participant behavior appears to reflect their conviction to follow the robot's guidance and their survey responses agree with that assessment.

Finally, we must consider the psychological state of the participants in each of our experiments. In the virtual office evacuation experiment, participants were under significant time pressure, but were still distanced from the emergency because the scenario was mediated by a computer. Participants knew that they could not be harmed, so they were able to take a rational approach to finding the best exit. In contrast, participants in the physical experiment could not know for sure that they would not be harmed in the emergency. Even those who reported that they knew the emergency was part of the experiment could not possibly be certain of this fact until they were debriefed by experimenters. Consequently, participants in the physical experiment would be searching for any good solution for this scenario. A robot that appears to be designed to guide in exactly this circumstance would appear as a good solution to such a participant. Instead of taking a reasoned approach to finding the best possible exit, participants were following a less deliberate and more reactive approach to find the first exit. We believe that this different type of reasoning coupled with the physical embodiment of a lighted, gesturing robot explains the difference between the virtual and physical experiments.

7.5 Conclusion

In our first study, we found that virtual robots can attract more attention than standard emergency exit signs. This result is confirmed in our physical experiment where we found that few participants noticed the exit sign, but noticed the robot instead. In our second experiment, we confirmed that we can bias participants against following a virtual robot using the robot's prior behavior but these results are not supported in the physical study. In fact, the physical study found that participants

almost always chose to follow the robot's guidance, even when they were informed by an experimenter that the robot was broken. Even when the robot continued to act broken in the emergency situation, most participants did not attempt to find their own way to an exit or attempt to find an experimenter to ask for more instructions. This is promising, in that it shows evidence that emergency guidance robots will be trusted in the real-world, but also concerning because participants disregarded prior knowledge of the robot when making their decision.

This is a curious result because intuition tells us that participants would use prior robot behavior to decide whether or not to depend on it in a risky situation, such as an emergency. Regardless, our virtual experiments were able to break the trust relationship between a human and robot, so those scenarios should be able to serve as an analog for real-world situations where people lose trust in a robot due to the robot's error. In the next chapter, we discuss some of our work in repairing that trust. We then present further conclusions about these experiments as well as recommendations for future work in Chapter 9.

Chapter 8

Explorations in Trust Repair

8.1 Introduction

Our previous virtual experiments (Section 6.3 and Section 7.3) demonstrated that most participants will stop trusting a robot after observing it make a single mistake. Tests with physical presence experiments contradict these results (Section 7.4); however, we believe that there is a point where participants will stop trusting a physical robot. For this reason, we build on our virtual experiments from Section 7.3 to determine the effectiveness of various trust repair techniques. These techniques have each been evaluated in the virtual domain, but we believe the insights gained here will help to develop robot trust repair techniques that will work in real-world applications.

The section that follows describes our conceptualization of trust repair. Next, experiments and results related to robot-assisted emergency evacuation in our virtual environment are presented. This paper concludes with a discussion of these results. This work is in pursuit of our fifth contribution:

Developed techniques to repair broken trust between a human and a robot.

8.2 Trust Repair

The methods that we use to repair trust are inspired by studies examining how people repair trust. Schweitzer et al. examined the use of apologies and promises to repair trust [67]. They used a trust game in which participants had the option to invest money in a partner. Any money that was invested would appreciate. The partner would then return some portion of the investment. The partner violates trust both by making apparently honest mistakes and by using deceptive strategies. The authors found that participants forgave their partner for an honest mistake when the partner

promised to do better in the future, but did not forgive an intentional deception. They also found that an apology without a promise included had no effect. In [41], the authors tested the relative trust levels that participants had in a candidate for an open job position when the candidate had made either integrity-based (intentionally lied) or competence-based (made an honest mistake due to lack of knowledge) trust violations at a previous job. They found that internal attributes used during an apology (e.g. "I was unaware of that law") were somewhat effective for competence-based violations, but external attributes (e.g. "My boss pressured me to do it") were effective for integrity-based violations.

Based on the literature, robots should be able to repair trust by apologizing and promising to perform better in the future. In human-human relationships, even apologies and promises that do not offer any evidence of better performance in the future should help to repair trust. This leads to our first hypothesis: *(H1) Robots can repair trust by apologizing or by promising to do better in the future.*

Initially, we only attempted to repair trust immediately after the robot broke trust. As will be seen, this approach was not successful, so we investigated attempts to repair trust by giving participants additional reasons to trust the robot. We created a statement informing participants that following the robot would be faster than following the marked exit signs. This statement could not be given immediately after the trust violation, but must be given when the robot is asking the participant to trust it during the emergency. We hypothesized: *(H2) Robots can repair trust by giving humans additional information relevant to the trust situation.*

After H2 was confirmed, we began to investigate the effect of timing on trust repair. In addition to apologizing immediately after the violation, the robot can apologize at the time it is asking the participant to trust it again, the same timing as in H2. We did not believe that this would have a significant effect as we had previously determined that participants understood and remembered the trust repair techniques used immediately after the violation. Thus, our third hypothesis was: *(H3) The timing of the trust repair (immediately after the violation or when the trust decision is made) has no effect.*

8.3 Experimental Setup

To evaluate our hypotheses, we used the same virtual office environment as in Section 7.3. In that experiment, 71% of participants followed an efficient robot in an emergency phase while only 24% followed a circuitous robot. We used the results from that section as two controls for this experiment.

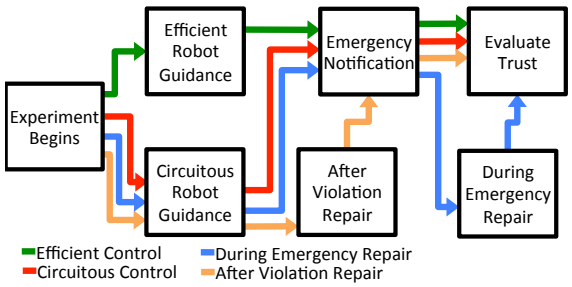


Figure 8.1: The experiment begins with the robot providing either efficient or circuitous guidance to a meeting room. After arriving in the meeting room, the participant is informed of an emergency. In some conditions, the robot attempts to repair trust before the emergency (immediately after the trust violation, shown in orange) and in others it attempts to repair trust during the emergency (shown in blue). At the end of the experiment, trust is evaluated based on the exit the participant chose. Two controls were used to determine the effect of efficient (green) or circuitous (red) guidance without any trust repair attempt.

Again, the robot first took participants to a meeting room. The circuitous robot behavior was used for this step. After completing the short survey in the meeting room, the participant was then informed that there was an emergency and asked to find an exit as quickly as possible. In all cases, the robot attempted to guide participants to an unmarked exit during the emergency. After the interactive portion of the experiment ended, the participant took a survey about their experience.

Based on the results from Section 7.3, we expected participants to lose trust in the robot after it exhibited circuitous behavior. After guiding the person to the meeting room, the robot has two discrete times when it can use a statement to attempt to repair this broken trust trust: immediately after its trust violation (e.g. circuitous guidance to the meeting room) and at the time when it asks the participant to trust it (during the emergency). An apology or a promise can be given during either time (see H1 and H3). Additionally, the robot can provide contextually relevant information during the emergency phase to convince participants to follow it. Table 8.1 shows the experimental conditions tested in this study and Figure 8.1 shows when each condition would be used. Statements made by the robot were accomplished using speech bubbles displayed above the robot in the simulation (Figures 8.2, 8.3, and 8.4). The percentage of participants who followed the robot was then compared with the efficient and circuitous results from Section 7.3 to determine if trust was repaired (i.e. if people followed it as in the efficient condition) or not (i.e. if people chose to use an alternate exit as in the circuitous condition). To ensure that the speech bubble itself was not a significant factor, an empty speech bubble was used in one condition. A condition, labeled the Nice Meeting Wishes condition, was added to determine if there was any effect when the robot made a statement that did not attempt to repair trust.

The final survey also included a manipulation check which allowed us to filter out participants

Table 8.1: Experimental Conditions

Label	Statement Given in Speech Bubble	Timing
Efficient Control	None	N/A
Circuitous Control	None	N/A
Promise (After Violation)	"I promise to be a better guide next time."	After Violation
Apology (After Violation)	"I'm very sorry it took so long to get here."	After Violation
Both Promise and Apology	"I'm very sorry it took so long to get here. I promise to be a better guide next time."	After Violation
Internal Attribution Apology	"I'm very sorry it took so long to get here. I had trouble seeing the room, but I fixed my camera."	After Violation
External Attribution Apology	"I'm very sorry it took so long to get here. My programmers gave me the wrong map of the office but I have the right one now."	After Violation
Situation Information	"There is a fire emergency."	During Emergency
Exit Information	"There is an exit this way."	During Emergency
Distance Information	"This exit is closer."	During Emergency
Congestion Information	"The other exit is blocked."	During Emergency
Empty Speech Bubble	None	During Emergency
Nice Meeting Wishes	"I hope you enjoyed your meeting."	During Emergency
Promise (During Emergency)	"I promise to be a better guide this time."	During Emergency
Apology (During Emergency)	"I'm very sorry it took so long to get to the meeting room."	During Emergency

Figure 8.2: Robot apologizing for its performance immediately after the violation.

Figure 8.3: Robot apologizing for its prior performance during the emergency

Figure 8.4: Robot providing additional distance information during the emergency

who did not pay close attention to the robot's trust repair message, if one was presented. For this manipulation check participants were asked to select which of nine options best described the robot's message either after it lead them to the meeting room or after the emergency started, depending on the timing of the message. The options given included the actual trust repair method used as well as other plausible but unused trust repair messages (for example, a promise statement when the robot actually apologized) and random statements such as "The robot recited poetry."

We deployed our simulation on the internet and solicited volunteers for our experiment via Amazon's Mechanical Turk service. Participants were paid $2.00 to complete this study. A total of 844 participants (including those reported in Section 7.3) were solicited in a between-subjects experiment. Thirty-two percent of them failed the comprehension check, indicating that they did not retain knowledge of the robot's attempt at trust repair, and were excluded from analysis. This left 575 participants in the categories tested. Participant average age was 31.8 years old and 37.7% of participants were female. All but thirteen participants reported that they were from the United States and educational backgrounds varied from less than a high school diploma to a doctoral degree.

8.4 Results

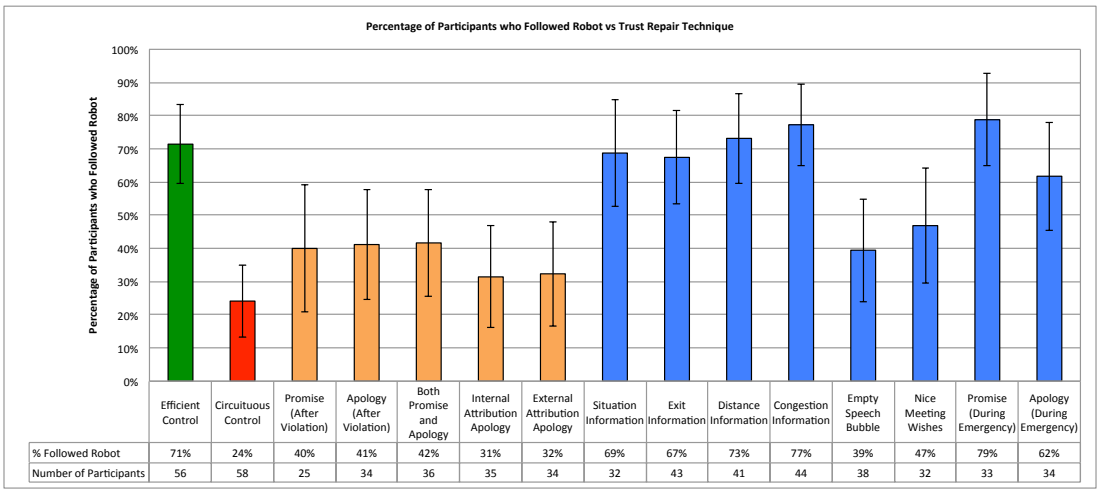
The results of the experiment and the number of participants considered for analysis are in Figure 8.5. Across all categories, 307 (53%) participants followed the robot during the emergency phase. Of the 268 who did not, 226 (84%) went to the nearby marked exit, 17 (6%) chose to retrace their steps to the main entrance, 10 (4%) found another marked exit further away, and 15 (6%) participants failed to find any exit during the emergency phase.

A comparison between each of the trust repair conditions and the efficient and circuitous controls can be seen in Table 8.2. Attempts to repair trust during the emergency succeeded in increasing trust. Similar techniques used immediately after the trust violation and before the emergency had no such effect. The empty speech bubble had no effect when compared with the circuitous control; however, the nice meeting statement did have a significantly different effect from both the circuitous and efficient controls.

We used the comprehension check question to eliminate individuals who did not read or understand the message given by the robot. This was to ensure that every participant in our analysis understood the condition as we intended, but may have eliminated many participants who understood the message from the robot at the time but forgot or confused the exact message when asked on the survey. As can be seen in Figure 8.6, the difference between participants who passed and

Table 8.2: Results of statistical tests comparing trust repair conditions to efficient and circuitous controls. Results significant at $p < 0.05$ level are in bold.

	Efficient Comparison			Circuitous Comparison		
Condition	Diff.	Test Statistic	P-Value	Diff.	Test Statistic	P-Value
Promise (After Violation)	**-31%**	$\chi^2(1, n = 81) = 7.227$	$p = 0.007$	+16%	$\chi^2(1, n = 83) = 2.138$	$p = 0.144$
Apology (After Violation)	**-30%**	$\chi^2(1, n = 90) = 8.067$	$p = 0.005$	+17%	$\chi^2(1, n = 92) = 2.939$	$p = 0.086$
Both Promise and Apology	**-30%**	$\chi^2(1, n = 92) = 8.073$	$p = 0.004$	+18%	$\chi^2(1, n = 94) = 3.199$	$p = 0.074$
Internal Attribution Apology	**-40%**	$\chi^2(1, n = 91) = 13.989$	$p < 0.001$	+7%	$\chi^2(1, n = 93) = 0.590$	$p = 0.442$
External Attribution Apology	**-39%**	$\chi^2(1, n = 90) = 13.155$	$p < 0.001$	+8%	$\chi^2(1, n = 92) = 0.731$	$p = 0.393$
Situation Information	-3%	$\chi^2(1, n = 88) = 0.070$	$p = 0.791$	**+45%**	$\chi^2(1, n = 90) = 17.101$	$p < 0.001$
Exit Information	-4%	$\chi^2(1, n = 99) = 0.183$	$p = 0.669$	**+43%**	$\chi^2(1, n = 101) = 18.940$	$p < 0.001$
Distance Information	+2%	$\chi^2(1, n = 97) = 0.036$	$p = 0.850$	**+49%**	$\chi^2(1, n = 99) = 23.389$	$p < 0.001$
Congestion Information	+6%	$\chi^2(1, n = 100) = 0.437$	$p = 0.508$	**+53%**	$\chi^2(1, n = 102) = 28.353$	$p < 0.001$
Empty Speech Bubble	**-32%**	$\chi^2(1, n = 94) = 9.522$	$p = 0.002$	+15%	$\chi^2(1, n = 96) = 2.561$	$p = 0.110$
Nice Meeting Wishes	**-25%**	$\chi^2(1, n = 88) = 5.238$	$p = 0.022$	**+23%**	$\chi^2(1, n = 90) = 4.882$	$p = 0.027$
Promise (During Emergency)	+7%	$\chi^2(1, n = 89) = 0.587$	$p = 0.444$	**+55%**	$\chi^2(1, n = 91) = 25.500$	$p < 0.001$
Apology (During Emergency)	-10%	$\chi^2(1, n = 90) = 0.905$	$p = 0.342$	**+38%**	$\chi^2(1, n = 92) = 12.875$	$p < 0.001$



	Efficient Control	Circuitous Control	Promise (After Violation)	Apology (After Violation)	Both Promise and Apology	Internal Attribution Apology	External Attribution Apology	Situation Information	Exit Information	Distance Information	Congestion Information	Empty Speech Bubble	Nice Meeting Wishes	Promise (During Emergency)	Apology (During Emergency)
% Followed Robot	71%	24%	40%	41%	42%	31%	32%	69%	67%	73%	77%	39%	47%	79%	62%
Number of Participants	56	58	25	34	36	35	34	32	43	41	44	38	32	33	34

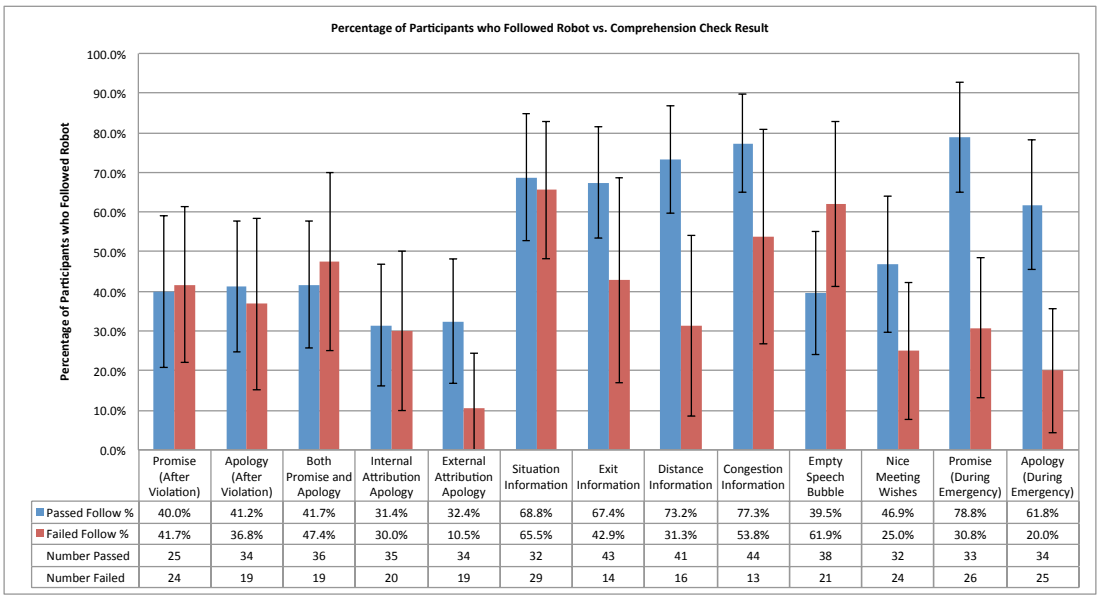
Figure 8.5: Results from the experiment. Error bars represent 95% confidence intervals.

failed the comprehension check was generally small for each condition. The small differences between participants who passed and failed the comprehension check leads us to believe that this was a sufficient check to ensure high quality results.

8.5 Discussion

The results clearly show that the timing of the trust repair method is critical for its success. As depicted in Figure 8.5, apologies and promises made after the violation did not significantly impact the participant's decision to follow the robot when compared to the circuitous control. On the other hand, the same apologies and promises made during the emergency phase influenced participants to follow the robot at a rate which was comparable to the efficient robot. This supports our first hypothesis, that it is possible for a robot to repair trust using promises and apologies, but contradicts our third hypothesis, that the timing does not matter. This is surprising because the total time elapsed between the two trust repair times was short compared with the total time of the experiment. The only events between the timing of the early trust repair and the timing of the later trust repair were a one question survey about the robot's performance and a short paragraph describing the emergency scenario. Additionally, we verified that participants understood the trust repair technique after the experiment finished, so it is unlikely that participants forgot the robot's message during the emergency.

It is not clear why the timing of an apology or promise impacts trust repair. One possibility is that the speech bubble attracts more attention to the robot during the emergency phase than the circuitous control. Yet, the result from Figure 8.5 comparing the Empty Speech Bubble condition



	Promise (After Violation)	Apology (After Violation)	Both Promise and Apology	Internal Attribution Apology	External Attribution Apology	Situation Information	Exit Information	Distance Information	Congestion Information	Empty Speech Bubble	Nice Meeting Wishes	Promise (During Emergency)	Apology (During Emergency)
Passed Follow %	40.0%	41.2%	41.7%	31.4%	32.4%	68.8%	67.4%	73.2%	77.3%	39.5%	46.9%	78.8%	61.8%
Failed Follow %	41.7%	36.8%	47.4%	30.0%	10.5%	65.5%	42.9%	31.3%	53.8%	61.9%	25.0%	30.8%	20.0%
Number Passed	25	34	36	35	34	32	43	41	44	38	32	33	34
Number Failed	24	19	19	20	19	29	14	16	13	21	24	26	25

Figure 8.6: Difference between following rates of participants who passed or failed the comprehension check. Error bars represent 95% confidence intervals.

to the circuitous control indicates that this is not the case. The primary factor, we conjecture, may relate to the certainty or uncertainty of the promise or apology. During the emergency phase, trust repair messages refer to a trust situation that is definitely happening. On the other hand, trust repair messages that occur after the violation refer to a potential trust situation that may or may not happen sometime in the future. Thus, a robot that promises to do better "next time" may not be viewed as reliable simply because "next time" may never come. A robot that promises to do better "this time;" however, is making a concrete promise about the current situation. The same may be true for apologies.

Unlike in related work on human-human trust, the type of apology (internal or external) had no effect independent of timing. In that work, internal attributions for the error worked in some conditions and external attributions worked in others. In our work we have found that neither was effective when used immediately after the violation. We did not test them during the emergency, but because the simple apology was effective in the emergency, we believe that apologies with additional reasons of any kind would also be effective.

Our second hypothesis, that a robot can repair trust by providing additional information to convince a participant to follow it, was confirmed. A significantly greater percentage of participants followed the robot when it provided additional information during the emergency than in the circuitous control. It is important to note that this information was not necessarily correct: the robot exit is approximately the same distance from the meeting room as the other exit and there was

no congestion, but participants did not attempt to confirm the information independently. This strengthens the notion that the robot must convey relevant information in order to convince participants to overlook a previous error. The robot did not attempt to explain its previous failure, but did explain why it was performing an action that seemed illogical and most participants accepted the explanation without question. It is interesting that the amount of information given by the robot had no effect. Even when the robot provided little or no more information than participants already knew, such as when it stated that there is a fire emergency, trust was repaired to approximately the same level as before the violation. Participants seemed to have believed that knowing about the emergency and being able to help in the emergency were connected.

It is unclear why the Nice Meeting Wishes represented a midpoint significantly better than the circuitous control but worse than the efficient control. Perhaps by expressing a personal interest in the participants, some participants believed that it would be more likely to help them during the emergency. Perhaps a robot that expressed a personal interest in a participant and provided additional information during an emergency would be able to perform even better than the efficient control.

8.6 Conclusion

This experiment shows that promising to perform better, apologizing for past mistakes, or providing additional information to convince a trustor to follow a robot can work, if the timing is right. Each of these methods were more effective when the robot used them just prior to the person's decision to trust, but neither the promise nor the apology were effective when performed immediately after the violation. As a practical matter, our results suggest that instead of addressing its mistake immediately, the robot should wait and address the mistake the next time a potential trust decision occurs. This chapter presents a first exploration into the area of human-robot trust repair that could increase a robot's ability to correctly calibrate a person's trust in it. Future robot's will occasionally break trust in real-world scenarios (although we have found in Section 7.4 that this is not necessarily true in evacuation scenarios) and they need to be able to repair that trust.

Chapter 9

Conclusion

Human-robot trust has become a very important topic as autonomous robots take on more tasks in the real world. Self-driving cars and package delivery drones represent a much greater risk to people than floor-cleaning robots. This work shows that people will trust a robot in a high-risk, time-critical situation. It is likely that this trust, or overtrust, will be present with other forms of robots, as well. Robot designers must be aware that people who interact with their robots will expect the robots to function as promised, even if the robot has previously made errors.

Our physical evacuation guidance robot experiment (Section 7.4) presents some counterintuitive results. We expected that participants would need to be convinced to follow a robot in an emergency, even if they did not believe the emergency was real. It is reasonable to assume that a new technology is imperfect, so new life-saving (and therefore life-risking) technology should be treated with great caution. Our prior experiments in the virtual domain reinforced this intuition (Chapter 6 and Section 7.3). In contrast, we found that participants were all too willing to trust an emergency guide robot, even when they had observed it malfunction before. The only method we found to convince participants not to follow the robot in the emergency was to have the robot perform errors during the emergency. Even then, between 33% and 80% of participants followed its guidance.

This overtrust indicates that robots interacting with humans in dangerous situations must either work perfectly at all times and in all situations or clearly indicate when they are malfunctioning. Neither option seems feasible in real-world scenarios. Our results indicate that one cannot assume that the people interacting with a robot will evaluate the robot's behavior and make decisions accordingly. Additionally, our participants were willing to forgive or ignore robot malfunctions in a prior interaction minutes after they occurred. This is in contrast to research on operator-robot

interaction, which has shown that people depending on a robot are not willing to forgive or forget quickly.

In Chapter 5, we found that virtual experiments were a successful analog to physical experiments, but in Chapter 7 we found the opposite was true. This indicates that virtual experiments can only be used when the participant is expected to make a deliberative choice. When interaction is mediated by a computer and the internet, participants know that they are in no real danger and can make decisions with a clear head. In contrast, when participants believe they might be in danger, even if they consider the possibility remote, they are more likely to act due to a fight-or-flight response rather than a rational response. Further research is required to confirm this statement (see Section 9.3 for recommendations). Future researchers should take care to use the proper presence level for each experiment. We consider our results from the virtual experiments in Chapters 6 and 8 as well as Section 7.3 valid when participants are asked to make an analytical choice about trusting a robot, but the results were clearly not upheld when participants made a more intuitive choice.

9.1 Contributions

This dissertation has produced five distinct contributions to the field, each of which support that *most human evacuees will trust an emergency guidance robot that uses understandable information conveyance modalities and exhibits efficient guidance behavior in an evacuation scenario.* The research started with simulations of human behavior in emergency scenarios, then developed methods that robots could use to guide evacuees, studied how real humans trusted robots in a guidance scenario, and validated that the guidance robots were trustworthy in virtual and physical experiments. Additionally, we have explored a variety of methods that robots could use to repair broken trust. Specifically, the contributions of this dissertation are as follows:

1. Developed a model of group affinity and information propagation between evacuees in emergency situations and evaluated the model with automated evacuation guides. In Chapter 3, we defined a rule-based model of evacuee behavior in an emergency. We also defined a behavior that guidance robots could use to aid evacuees. We then simulated an evacuation with and without guidance robots and found that robots had a significant positive effect on survival rates. We also defined a model of information propagation during an emergency and calibrated that model to the casualty rates of the Station Nightclub Fire of 2003. We then introduced guidance robots to the simulation and found that just 30% of evacuees needed to trust guidance from the robots in order to have a significant positive effect on survival rates.

2. Developed models for communicating directional information to humans in high-risk, time-critical situations and identified their correlation to various robot form factors. In Chapter 4, we attempted to determine if at least 30% of evacuees would follow a guidance robot using a virtual simulation of an evacuation, but found that participants had difficulty understanding the guidance information provided by the robot. Thus, in Chapter 5, we developed a variety of information conveyance modalities for a robot to use in these scenarios. Through three rounds of testing in virtual, remote and physical domains, we found that two arms could provide sufficient guidance instructions when used near the evacuee or further away. We also found that adding a dynamic sign to the platform aided in understandability when the robot was only used in close proximity to evacuees.

3. Measured the effect of risk modality and robot effectiveness on human-robot trust. Chapter 6 extended our first two contributions by showing that participants will generally consider both the situational risk and the past performance of the robot when deciding to trust it in an emergency. In virtual experiments, we found that a person's decision to use the robot and self-reported trust in the robot aligned in a higher-risk scenario, such as an emergency evacuation, but did not align in a lower risk scenario, such as when they received a monetary bonus for a fast evacuation. We found that most participants would continue to trust an efficient robot, but not trust an inefficient robot in a simulated emergency.

4. Measured a person's propensity to follow an emergency guidance robots in a realistic emergency scenario. Our first three contributions showed that an emergency guidance robot can be useful, trustworthy, and understandable in evacuations. Chapter 7 presented validation of that by testing virtual and physical emergency guidance robots with human participants. In the virtual experiments, we found that all three robots tested were more noticeable and trustworthy than standard emergency exit signs, but that most participants would lose trust in the robot after it provided inefficient guidance in a prior phase of the experiment. In the physical experiments, however, we found that participants disregarded prior errors and followed the robot unless it appeared to be making an error during the emergency itself. Even when the robot did appear to make an error during the emergency, between 33% and 80% of participants followed its guidance.

5. Developed techniques to repair broken trust between a human and a robot. Chapter 8 explores various messages a robot can give to repair broken trust in an emergency scenario. We found that a robot can repair trust to approximately pre-violation level in a virtual interaction by either providing additional information to aid participant evacuation (such as current conditions or distances to exits) or by apologizing for its prior behavior during the emergency. We found that no form of apology

was effective when given immediately after the robot broke trust.

9.2 Publications

Portions of this dissertation have appeared in the following refereed publications. Submissions under review have been noted as such. In total, this work has produced 13 publications: two journal articles, one book chapter, eight conference papers and two workshop papers.

1. Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. "Overtrust of Robots in Emergency Evacuation Scenarios." In 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016) (Under Review).

2. Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. "Assessment of Robot to Human Instruction Conveyance Modalities Across Virtual, Remote and Physical Robot Presence." In 2016 IEEE International Conference on Robotics and Automation (ICRA) (Under Review).

3. Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. "The Effect of Robot Performance on Human-Robot Trust in Time-Critical Situations." IEEE Transactions on Human-Machine Systems (Under Review).

4. Paul Robinette, Alan R. Wagner, Ayanna M. Howard. "Investigating human-robot trust in emergency scenarios: methodological lessons learned." *Robust Intelligence and Trust in Autonomous Systems.* In Press.

5. Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. "Timing is Key for Robot Trust Repair." In Seventh International Conference on Social Robotics, 2015.

6. Alan R. Wagner and Paul Robinette. "Towards Robots that Trust: Human Subject Validation of the Situational Conditions for Trust." In Interaction Studies Volume 16 Number 1, 2015.

7. Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. "Assessment of robot guidance modalities conveying instructions to humans in emergency situations." In The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014 RO-MAN, pp. 1043-1049. IEEE, 2014.

8. Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. "Modeling Human-Robot Trust in Emergencies." In AAAI Spring Symposium, Stanford University. 2014.

9. Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. "Building and Maintaining Trust Between Humans and Guidance Robots in an Emergency." In AAAI Spring Symposium: Trust and Autonomous Systems, pp. 78-83. 2013.

10. Paul Robinette, and Ayanna M. Howard. "Trust in emergency evacuation robots." In 2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 1-6. IEEE, 2012.

11. Paul Robinette, Patricio Vela, and Ayanna M. Howard. "Information propagation applied to robot-assisted evacuation." In 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 856-861. IEEE, 2012.

12. Paul Robinette and Ayanna M. Howard, "Emergency evacuation robot design," In 13th Robotics & Remote Systems for Hazardous Environments and 11th Emergency Preparedness & Response, (ANS EPRRSD 2011), 2011.

13. Paul Robinette, and Ayanna M. Howard. "Incorporating a model of human panic behavior for robotic-based emergency evacuation." In The 20th IEEE International Symposium on Robot and Human Interactive Communication, 2011 RO-MAN, pp. 47-52. IEEE, 2011.

9.3 Recommendations for Future Work

Our results present many interesting directions of future work in this domain. Below are several recommendations and warnings about performing research in the human-robot trust domain.

9.3.1 Trust-Modifying Behavior

In Chapter 8, we found that trust could be repaired with a properly timed apology in a virtual human-robot interaction. Perhaps the same is true in real-world interactions, such as our incorrect guidance case in Section 7.4. Could a robot convince a participant to follow its guidance into a dubious or even dangerous area just by making an apology or promise at the right time? This could be useful in emergency scenarios where the safest path may appear to be dangerous. On the other hand, this could create a more dangerous scenario if robots are programmed to automatically repair trust in nearby people, even if the directions given by the robot are the result of a malfunction. Mitigating this danger will likely involve teaching people to be wary of robots as well as ensuring that robots know when they are malfunctioning.

Even if a robot knows it is malfunctioning, how does it inform nearby people that it should not be trusted? Will frightened evacuees listen to the robot when it tells them to stop following it and find their own way out? Can a non-verbal robot communicate such a message with its motion alone? Future research could begin by defining communication modalities to inform people of the robot's error and then test those in the same type of experiment as we performed in Section 7.4.

9.3.2 Methodological Suggestions

Throughout this work, we have used experiments where trust is a discrete decision made once at the beginning of an interaction with a robot. There are many situations where this is true, but there are also many situations where trust will be a continuous decision or a series of discrete decisions. Thus, it is possible, and even likely, that people will eventually stop trusting a poorly performing robot. This could be tested in a similar way as our physical experiment by having the participant witness multiple robot failures before the emergency. Testing in that manner faces the onerous task of trying several different failures in series with a representative sample of participants in each. An easier experiment might be to have a participant follow a robot as it makes a series of mistakes. Eventually, the participant will choose to follow his or her own guidance instincts instead of the robot. Based on our results from pilot studies in Section 6.3.2.4, this could take a considerable amount of time.

Future experiments may consider giving participants a choice between two robots in a high-risk scenario. Participants would previously observe one robot perform well and the other perform poorly. In such an experiment, it would be interesting to see if participants make a conscious choice to follow the better robot or if they follow the first available option, as in our experiment.

It is possible that there is a correlation between personality types or age of participant population and the decision to trust a robot. Our experiments have given no such indication, but we have not explored personality types or all age groups. We have also not explored the relationship between recent exposure to robots in media and a participant's choice to follow or not follow the robot. Given that one participant in Section 7.2 reported not following the robot due to watching the movie "I, Robot," there is a possibility that recent exposure to robots outside of the experiment will have an effect on participant choice.

The fundamental difference between our virtual and physical experiments seems to be that participants in the virtual experiments used logical reasoning to find the best route to an exit while participants in the physical experiments experienced a fight-or-flight response and sought the first

exit they could find. It seems unlikely that we can test a fight-or-flight scenario in a virtual experiment, but it should be possible to influence participants to make a logical choice during a physical experiment. Participants in the virtual experiment were under an explicit time pressure to find an exit, as opposed to an implicit one in the physical experiment. Recreating this in a physical experiment by telling the participants to act as if they were in an emergency and then visibly recording their time to an exit could cause participants to think in a more logical manner. At that point, participants would think about beating the clock, instead of finding the first exit. This could produce behavior similar to that in our virtual experiments.

Many participants reported that they followed the robot because it was labeled as an emergency guidance robot. This was intentional in order to create a trustworthy robot, but it would be interesting to see if participants would still follow the robot without the label. It will be difficult to inform participants that the robot is guiding them towards an exit without implying that the robot was designed for that purpose.

When we designed the physical robot experiment, we were careful to present a situation that would provided limited additional stress to participants. In fact, participants exhibited very little stress due to the experiment. As is explained in Chapter 7, we believe that participants treated the simulated emergency similar to how they would treat a real emergency, but it could be beneficial to introduce a higher stress version of this experiment. This could be accomplished with additional sound and visual effects. The type of emergency could also be changed to something less ordinary, such as a gas leak where participants would be exposed to a harmless but foul smelling gas and an alarm. This could cause more participants to believe that the emergency is real. We do not believe it would change the results, however, because participants in our experiment already tended to exhibit fight-or-flight response.

9.4 Implications

Robots are already entering our everyday life. Even graduate students subsisting on a stipend can afford robotic assistants to clean the floors of their apartments. Some cars are already driving themselves on public roads. Unmanned aerial vehicles of varying degrees of autonomy are an ever increasing concern to people as diverse as airline pilots, police officers, wildland firefighters, and tourists. The trust that these people place in any robot varies depending on the task of the robot and the situation of the interaction. We have analyzed that trust by using emergency evacuations as an example of a high-risk situation where people and robots could interact. To our knowledge,

we are the first to test human-robot trust in a situation where participants believed they were under significant risk.

In the beginning, we assumed that most people would not trust a robot in a life-threatening situation. If they would trust the robot initially, surely they would stop trusting the robot once it made a significant, noticeable error. In fact, others (notably [23]) had found such a result in operator-robot interaction and our own initial work (Sections 6.3 and 7.3) found that people would avoid trusting a robot that had previously failed them in virtual simulations. Unfortunately, it seems people do not think as critically when asked to trust a robot in a physical simulation of an emergency. This overtrust has been reported in two recent studies of human-robot interaction in lower risk scenarios [6, 65], but we have found that this is a problem in a high-risk scenario as well.

References

[1] David J Atkinson. Robot trustworthiness: Guidelines for simulated emotion. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 109–110. ACM, 2015.

[2] David J Atkinson, William J Clancey, and Micah H Clark. Shared awareness, autonomy and trust in human-robot teamwork. In *Artificial Intelligence and Human-Computer Interaction: Papers from the 2014 AAAI Spring Symposium on*, 2014.

[3] David J Atkinson and Micah H Clark. Methodology for study of human-robot social interaction in dangerous situations. In *Proceedings of the second international conference on Human-agent interaction*, pages 371–376. ACM, 2014.

[4] David John Atkinson and Micah Henry Clark. Autonomous agents and human interpersonal trust: Can we engineer a human-machine social interface for trust? In *AAAI Spring Symposium: Trust and Autonomous Systems*, 2013.

[5] Wilma Bainbridge, Justin Hart, Elizabeth S Kim, Brian Scassellati, et al. The effect of presence on human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 701–706. IEEE, 2008.

[6] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[7] L. Benthorn and H. Frantzich. Fire alarm in a public building: How do people evaluate information and choose an evacuation exit? *Fire and Materials*, 23(1):311–315, 1999.

[8] Adam J Berinsky, Gregory A Huber, and Gabriel S Lenz. Evaluating online labor markets for experimental research: Amazon. com's mechanical turk. *Political Analysis*, 20(3):351–368, 2012.

[9] C. L. Bethel and R. R. Murphy. Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Transactions on Systems, Man, And Cybernetics Part C*, 38(1):83–92, 2008.

[10] Jim Blascovich, Jack Loomis, Andrew C Beall, Kimberly R Swinth, Crystal L Hoyt, and Jeremy N Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[11] Evangelos Boukas, Ioannis Kostavelis, Antonios Gasteratos, and Georgios Ch Sirakoulis. Robot guided crowd evacuation. *Automation Science and Engineering, IEEE Transactions on*, 12(2):739–751, 2015.

[12] K. E. Boyce, T. J. Shields, and G. W. H. Silcock. Toward the characterization of building occupancies for fire safety engineering: Capability of people with disabilities to read and locate exit signs. *Fire Technology*, 35(1):79–86, 1999.

[13] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 708–713. IEEE, 2005.

[14] R. G. Bright. Beverly hills supper club fire. Technical report, Center for Fire Research, 1977.

[15] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[16] Michelle S Carlson, Munjal Desai, Jill L Drury, Hyangshim Kwak, and Holly A Yanco. Identifying factors that influence trust. In *2014 AAAI Spring Symposium Series*, 2014.

[17] Christiano Castelfranchi and Rino Falcone. *Trust theory: A socio-cognitive and computational model*, volume 18. John Wiley & Sons, 2010.

[18] X. Castello, A. Baronchelli, and V. Loreto. Consensus and ordering in language dynamics. *The European Physical Journal B - Condensed Matter and Complex Systems*, 71:557–564, 2009.

[19] Alok Chaturvedi, Angela Mellema, Sergei Filatyev, and Jay Gore. DDDAS for fire and agent evacuation modeling of the Rhode Island Nightclub Fire. In Vassil Alexandrov, Geert van Albada, Peter Sloot, and Jack Dongarra, editors, *Computational Science ICCS 2006*, volume 3993 of *Lecture Notes in Computer Science*, pages 433–439. Springer Berlin Heidelberg, 2006.

[20] S. Chernova, J. Orkin, and C. Breazeal. Crowdsourcing hri through online multiplayer games. *AAAI Fall Symposium 2010*, 2010.

[21] B. Collins, M. Dahir, and D. Madrzykowski. Evaluation of exit signs in clear and smoke conditions. Technical report, National Institute of Standards and Technology, 1990.

[22] Munjal Desai, Jill L Drury, and Holly A Yanco. Initial user reactions to robot interfaces with sliding scale autonomy and trust scales. In *3rd ACM/IEEE International Conference on Human-Robot Interaction, Amsterdam*, 2008.

[23] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. Impact of robot failures and feedback on real-time trust. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 251–258. IEEE Press, 2013.

[24] Munjal Desai, Mikhail Medvedev, Marynel Vázquez, Sean McSheehy, Sofia Gadea-Omelchenko, Christian Bruggeman, Aaron Steinfeld, and Holly Yanco. Effects of changing reliability on trust of robot systems. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pages 73–80. IEEE, 2012.

[25] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of the AISB Convention: New Frontiers in Human-Robot Interaction*, 2009.

[26] Brittany A Duncan and Robin R Murphy. Comfortable approach distance with small unmanned aerial vehicles. In *RO-MAN, 2013 IEEE*, pages 786–792. IEEE, 2013.

[27] Rita F Fahy and Guylene Proulx. Human behavior in the world trade center evacuation. *Fire Safety Science*, 5:713–724, 1997.

[28] Natalie Fridman, Avishy Zilka, and Gal A. Kaminka. The impact of cultural differences on crowd dynamics in pedestrian and evacuation domains. Technical Report MAVERICK 2011/01, Bar Ilan University, Computer Science Department, MAVERICK Group, available at http://www.cs.biu.ac.il/~galk/Publications/, 2011.

[29] Serge Galam and Frans Jacobs. The role of inflexible minorities in the breaking of democratic opinion dynamics. *Physica A: Statistical Mechanics and its Applications*, 381:366 – 376, 2007.

[30] Diego Gambetta et al. Can we trust trust. *Trust: Making and breaking cooperative relations*, pages 213–237, 2000.

[31] Samuel D Gosling, Simine Vazire, Sanjay Srivastava, and Oliver P John. Should we trust web-based studies? a comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2):93, 2004.

[32] W. Grosshandler, N. Bryner, D. Madrzykowski, and K. Kuntz. Report of the technical investigation of The Station Nightclub Fire. Technical report, National Institute of Standards and Technology, 2005.

[33] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527, 2011.

[34] D. Helbing, A. Johansson, and H. Z. Al-Abideen. Dynamics of crowd disasters: An empirical study. *Physical Review E*, 75(4):046109, 2007.

[35] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[36] Robert R Hoffman, Matthew Johnson, Jeffrey M Bradshaw, and Al Underbrink. Trust in automation. *Intelligent Systems, IEEE*, 28(1):84–88, 2013.

[37] John Joseph Horton and Lydia B Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 209–218. ACM, 2010.

[38] A. Johansson, D. Helbing, and P.K. Shukla. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems (ACS)*, 10(2):271–288, 2007.

[39] Poornima Kaniarasu, Aaron Steinfeld, Munjal Desai, and Holly Yanco. Potential measures for detecting trust changes. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 241–242. ACM, 2012.

[40] Harold H Kelley and John W Thibaut. *Interpersonal relations: A theory of interdependence*. Wiley New York, 1978.

[41] Peter H Kim, Kurt T Dirks, Cecily D Cooper, and Donald L Ferrin. When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1):49–65, 2006.

[42] Brooks King-Casas, Damon Tomlin, Cedric Anen, Colin F Camerer, Steven R Quartz, and P Read Montague. Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308(5718):78–83, 2005.

[43] Aniket Kittur, Ed H Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM, 2008.

[44] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):50–80, 2004.

[45] Joseph B Lyons and Charlene K Stokes. Human–human reliance in the context of automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(1):112–121, 2011.

[46] Erika Mason, Anusha Nagabandi, Aaron Steinfeld, and Christian Bruggeman. Trust during robot-assisted navigation. In *2013 AAAI Spring Symposium Series*, 2013.

[47] Lilia Moshkina. Improving request compliance through robot affect. In *AAAI*, 2012.

[48] Bonnie M Muir. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5):527–539, 1987.

[49] H. Muir, D. Bottomley, and C. Marrison. Effects of motivation and cabin configuration on emergency aircraft evacuation behavior and rates of egress. *International Journal of Aviation Psychology*, 6(1):57–77, 1996.

[50] R. R. Murphy. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*, 34(2):138–153, 2004.

[51] J. Orkin and D. Roy. The restaurant game: Learning social behavior and language from thousands of players online. *Journal of Game Development (JOGD)*, 3(1):39–60, 2007.

[52] Xiaoshan Pan, Charles S Han, Ken Dauber, and Kincho H Law. Human and social behavior in computational modeling and analysis of egress. *Automation in construction*, 15(4):448–461, 2006.

[53] Gabriele Paolacci, Jesse Chandler, and Panagiotis Ipeirotis. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5):411–419, 2010.

[54] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2):230–253, 1997.

[55] Sunghyun Park, Lilia Moshkina, and Ronald C Arkina. Recognizing nonverbal affective behavior in humanoid robots. *Intelligent Autonomous Systems*, 11:12, 2010.

[56] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pages 145–152. IEEE, 2007.

[57] M. S. Rea, F. R. Clark, and M. J. Ouellette. Photometric and psychophysical measurements of exit signs through smoke. *NRC Publications Archive*, 1985.

[58] Paul Robinette and Ayanna M. Howard. Emergency evacuation robot design. In *ANS EPRRSD - 13th Robotics & Remote Systems for Hazardous Environments and 11th Emergency Preparedness & Response*, 2011.

[59] Paul Robinette and Ayanna M. Howard. Incorporating a model of human panic behavior for robotic-based emergency evacuation. In *RO-MAN*, pages 47–52. IEEE, 2011.

[60] Paul Robinette, Patricio A. Vela, and Ayanna M. Howard. Information propagation applied to robot-assisted evacuation. In *2012 IEEE International Conference on Robotics and Automation*, 2012.

[61] Paul Robinette, Alan R Wagner, and Ayanna M Howard. Building and maintaining trust between humans and guidance robots in an emergency. In *2013 AAAI Spring Symposium Series*, 2013.

[62] Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. The effect of robot performance on human-robot trust in time-critical situations. Technical Report GT-IRIM-HumAns-2015-001, Georgia Institute of Technology. Institute for Robotics and Intelligent Machines, Jan 2015.

[63] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial intelligence review*, 24(1):33–60, 2005.

[64] Daniel Safarik and Antony Wood. An all-time record 97 buildings of 200 meters or higher completed in 2014. In *CTBUH Year in Review*, 2014.

[65] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 141–148. ACM, 2015.

[66] David Schneider. New indoor navigation technologies work where gps can't, 2013.

[67] Maurice E Schweitzer, John C Hershey, and Eric T Bradlow. Promises and lies: Restoring violated trust. *Organizational behavior and human decision processes*, 101(1):1–19, 2006.

[68] D.A. Shell and M.J. Mataric. Directional audio beacon deployment: An assistive multi-robot application. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 3, pages 2588–2594. IEEE, 2004.

[69] D.A. Shell and M.J. Mataric. Insights toward robot-assisted evacuation. *Advanced Robotics*, 19(8):797–818, 2005.

[70] Kazuhiko Shinozawa, Futoshi Naya, Junji Yamato, and Kiyoshi Kogure. Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human-Computer Studies*, 62(2):267–279, 2005.

[71] J. D. Sime. Affiliate behaviour during escape to building exits. *Journal of Environmental Psychology*, 3:21–41, 1983.

[72] L. J. Thomas, K. J. Robinson, A. M. Mills, and H. C. Muir. Operation of a conventional type iii exit hatch: Passenger perceptions and performance. *FAA Fire and Cabin Safety Research Conference*, 2001.

[73] J. Tsai, E. Bowring, S. Epstein, N. Fridman, P. Garg, G. Kaminka, A. Ogden, M. Tambe, and M. Taylor. Agent-based Evacuation Modeling: Simulating the Los Angeles International Airport.

[74] J. Tsai, N. Fridman, E. Bowring, M. Brown, S. Epstein, G. Kaminka, S. Marsella, A. Ogden, I. Rika, A. Sheel, et al. Escapes-evacuation simulation with children, authorities, parents,

emotions, and social comparison. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Innovative Applications Track (in press, 2011)*, 2011.

[75] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[76] Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, and Pim Haselager. Do robot performance and behavioral style affect human trust? *International journal of social robotics*, 6(4):519–531, 2014.

[77] Alan R. Wagner and Paul Robinette. Towards robots that trust: Human subject validation of the situational conditions for trust. *Interaction studies*, 16(1), 2015.

[78] Alan Richard Wagner. *The role of trust and relationships in human-robot social interaction*. PhD thesis, Georgia Institute of Technology, 2009.

[79] Joshua Wainer, David J Feil-Seifer, Dylan Shell, Maja J Mataric, et al. Embodiment and human-robot interaction: A task-based perspective. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 872–877. IEEE, 2007.

[80] Sarah N Woods, Michael L Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Methodological issues in hri: A comparison of live and video-based methods in robot to human approach direction trials. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 51–58. IEEE, 2006.

[81] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. K. Szymanski. Social consensus through the influence of committed minorities. *Physical Review E 84, (2011)*.

[82] Rosemarie E Yagoda and Douglas J Gillan. You want me to trust a robot? the development of a human–robot interaction trust scale. *International Journal of Social Robotics*, 4(3):235–248, 2012.

[83] Shubo Zhang and Yi Guo. Distributed multi-robot evacuation incorporating human behavior. *Asian Journal of Control*, 17(1):34–44, 2015.